Vol. 64, No. 3 May, 1957

Psychological Review

THEODORE M. NEWCOMB, Editor
University of Michigan

CONTENTS

On the Psychophysical Law.....S. S. STEVENS 153

Theory of Selective Learning With Probable Reinforcements.....FRANK RESTLE 182

Value and the Perceptual Judgment of Magnitude.....H. TAJFEL 192

A Mechanical Model for Human Attention and Immediate Memory.....D. E. BROADBENT 205

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Consulting Editors

Solomon E. Asch	Robert B. MacLeod
Robert R. Blake	David C. McClelland
Stuart W. Cook	George A. Miller
Clyde H. Coombs	Gardner Murphy
Leon Festinger	Oscar Oeser
W. R. Garner	Carl Pfaffmann
James J. Gibson	Carroll C. Pratt
D. O. Hebb	John P. Seward
Harry Helson	David Shakow
E. R. Hilgard	Richard L. Solomon
Carl I. Hovland	Eliot Stellar
E. Lowell Kelly	S. S. Stevens
David Krech	Eric Trist
Robert W. Leeper	Edward L. Walker
Kenneth MacCorquodale	Robert W. White

Arthur C. Hoffman, Managing Editor

Helen Orr, Assistant Managing Editor

Editorial Staff: Sadie J. Doyle, Frances H. Clark, Barbara Cummings

---

The *Psychological Review* is devoted to theoretical articles of significance to any area of psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding twelve printed pages (about 7,500 words) are not accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail. Manuscripts should be addressed to the Editor, Dr. Theodore M. Newcomb, Doctoral Program in Social Psychology, University of Michigan, Ann Arbor, Michigan.

PUBLISHED BIMONTHLY BY THE

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

PRINCE AND LEMON STREETS, LANCASTER, PA.

1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

$8.00 volume $1.50 issue

Entered as second-class matter July 13, 1897, at the post office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1948

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1957 by the American Psychological Association, Inc.

VOL. 64, NO. 3 MAY, 1957

THE PSYCHOLOGICAL REVIEW

ON THE PSYCHOPHYSICAL LAW[1]

S. S. STEVENS

Harvard University

After dealing his critics their deserts, Fechner concluded his polemic of 1877 with a defiant five-line *Nachwort* (10).

> The tower of Babel was never finished because the workers could not reach an understanding on how they should build it; my psychophysical edifice will stand because the workers will never agree on how to tear it down.

For nearly a century the critics have pecked at the edifice in the disorganized manner that inspired no fear in Fechner. They have contended that sensation is not measurable, and that it is; that Weber's law is not true, and that it is; that the just noticeable difference is not a proper unit, and that it is; that the logarithmic law is not found in experiment, and that it is. As James put it, Fechner's critics flailed about, "smiting his theories hip and thigh and leaving not a stick of them standing . . . ," but here we are a hundred years later still discussing Fechner.

The lesson of history is that a bold and plausible theory that fills a scientific need is seldom broken by the impact of contrary facts and arguments. Only with an alternative theory can we hope to displace a defective one.

The purpose here is to try to do just that—to try to show that there is a general psychophysical law relating subjective magnitude to stimulus magnitude, and that this law is simply that equal stimulus ratios produce equal subjective ratios. On numerous perceptual continua, direct assessments of subjective magnitude seem to bear an orderly relation to the magnitude of the stimulus. To a fair first-order approximation, the ratio scales constructed by "direct" methods (as opposed to the indirect procedures of Fechner) are related to the stimulus by a power function of one degree or another. Evidence for this fact is piling up under the impact of ratio-scaling procedures whose development over the past three decades has given a "new look" to psychophysics. The fruits of these new pursuits have implications and consequences for the traditional issues in the area of psychophysics, as well as for certain related issues concerned with category judgments, time-order errors, and scaling theory.

That the Fechnerian philosophy of indirect measurement will fade quietly away as soon as it has been shown what direct methods can achieve is scarcely to be counted on, for elaborations and applications of this philosophy pervade important segments of the psychological activity. Curiously enough, modern Fechnerians are found less frequently among the psychophysicists than among the psychometricians and scale construc-

[1] This research was supported partly by a grant from the National Science Foundation, partly by Contract Nonr-1866 (15) with the Office of Naval Research (Project Nr142-201, Report PNR-188). Reproduction for any purpose of the U. S. Government is permitted.

tors. It is they who sometimes assume that "equally often noticed differences are equal." As elaborated in Thurstone's "law of comparative judgment," this principle is essentially Fechner's principle: namely, that the unit of measurement is given by resolving power. If not explicitly, at least by implication, this philosophy of indirect measurement asserts that all we can know about magnitude is what confusion tells us. Variability becomes the measure of things, and the mean is meaningless. But on an important class of those simpler psychological continua where these notions are testable, we can show that equally often noticed differences are *not* equal, and that a scale proportional to psychological magnitude is not achieved by procedures that try to transform variabilities, discriminal dispersions, or confusions into units of measure.

Before we consider this point further we will need to draw a distinction between two classes of continua.

TWO CLASSES OF CONTINUA

Perceptual continua divide themselves into two general classes (**64**). The nature of this division is suggested in a general way by the traditional dichotomy between quantity and quality. Continua having to do with *how much* belong to what we have called Class I, or prothetic; continua having to do with *what kind* and *where* (position) belong to Class II, or metathetic. Class I seems to include, among other things, those continua on which discrimination is mediated by an additive or prothetic process at the physiological level (**53, 67**). An example is loudness, where we progress along the continuum by adding excitation to excitation. Class II includes continua on which discrimination is mediated by a physiological process that is substitutive, or metathetic. An example is pitch, where we progress along the continuum by substituting excitation for excitation, i.e., by changing the locus of excitation.

We know too little about physiological mechanisms to say whether all Class I continua are based on additive processes or all Class II continua on substitutive processes, but in those instances where the facts seem clear the parallels between function and physiology are at least suggestive. Until our knowledge stands improved, it is perhaps best to classify the perceptual continua by the pragmatic criteria of the way they behave and to hope that any uniformities we can discover will lead to deeper insights into basic mechanisms.

Four functional criteria are relevant to the distinction between prothetic and metathetic continua, although they have not all been tested with equal thoroughness. These four criteria concern the subjective size of the j.n.d., the form of category rating-scales, the time-order error, and hysteresis. It is likely, of course, that other criteria will be discovered if we continue to look for them.

1. *Subjective size of the j.n.d.* I list this criterion first because it is the growth in the subjective size of the j.n.d. as we go up the scale on a Class I or prothetic continuum that seems, in a sense, to "explain" at least two of the other three criteria. Fechner proposed that scales can be constructed by counting off just noticeable differences. The implication is that the sensation produced by a stimulus 50 j.n.d.'s above threshold is half as great as that produced by a stimulus 100 j.n.d.'s above threshold. The hard fact of the matter is that if the typical subject were confronted with two such stimuli on a Class I continuum he would assert with certainty that the ratio between the two sensations is greater than two, because scales obtained by summating j.n.d.'s are nonlinearly related to ratio scales of subjective magnitude (**33, 52**). But more about this in a later section. The

point here is that on Class I continua the j.n.d.'s are not equal in subjective size.

On continua of Class II (metathetic) the j.n.d.'s turn out to be approximately equal in subjective size when measured by magnitude scales of the continuum. The linearity between the mel scale of subjective pitch and the j.n.d. scale for frequency is a case in point (56) and it seems reasonable to suppose that the same linear relation is approximated on such continua as position and inclination.

2. *Category rating-scales.* A category rating-scale is the function obtained when a subject judges a set of stimuli in terms of a set of categories labeled either by numbers or by adjectives. The form of these scales is different on the two kinds of continua. As shown by studies of a dozen perceptual dimensions (64), the category scales on continua of Class I are concave downward when plotted against a ratio scale of the subjective magnitude. Category scales on continua of Class II *may be* linear when so plotted.

The chief factor that produces nonlinearity in the category scales of Class I is variation in the subject's sensitivity to differences. Near the lower end of the scale where discrimination is good the categories tend to be narrow, and by consequence the slope of the function is steep. Near the upper end, where a given stimulus difference is less easy to detect, the categories broaden and the slope declines. Only on Class II continua, where sensitivity (measured in subjective units) remains relatively constant, is it ordinarily possible to produce category scales that are linearly related to subjective magnitude.

Prothetic continua of Class I on which category scales have proved nonlinear include apparent length, area, numerousness, duration, heaviness, lightness, brightness and loudness. Metathetic

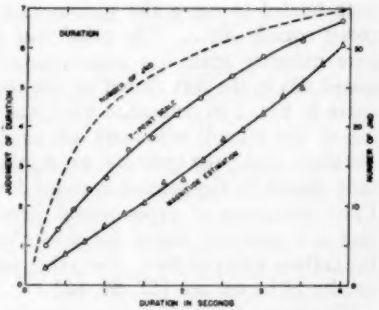


FIG. 1. J.n.d. scale, category scale, and ratio scale for apparent duration. *Triangles:* Mean judgments of 12 subjects who estimated the apparent duration of a white noise. Stimuli were presented in a different irregular order to each subject. *Circles:* Mean category judgments made by 16 subjects on a scale from 1 to 7. The end stimuli were presented at the outset to indicate the range, and each subject judged each duration twice in a random order. Stimulus spacing was adjusted to produce a "pure" category scale. *Dashed line:* Summated j.n.d.'s, right-hand ordinate.

continua of Class II on which category scales have proved more or less linear include visual position, inclination, proportion, and pitch. On both types of continua alterations in the form of the category scale may be produced by a variety of factors, including stimulus spacing (or relative frequency of presentation), landmarks, and differential familiarity.

By means of an iterative series of experiments we can achieve a stimulus spacing that will neutralize the effects of spacing and produce a "pure" category scale, uncontaminated by the subject's expectations regarding the frequency with which he should name the various categories. An example of an approximation to a pure category scale is shown in Fig. 1 (circles). This experiment was a follow-up to the experiments on duration reported earlier (64). Guided by the outcome of the earlier studies, I so spaced the stimuli that the 16 sub-

jects tended to name the various categories equally often. The proof that a pure category scale has been approximated lies in the fact that if we use the curve in Fig. 1 to determine what spacing of the stimuli will mark off equal distances along the ordinate, no significant change in the spacing is called for. (The technique of experimental iteration is a powerful device for neutralizing certain kinds of bias. For other examples of its use see [57, 64, 66].)

The iterated experiments on duration culminating in that shown in Fig. 1 converge on a category scale that is typical of those obtained on continua of Class I (prothetic). We note that it is concave downward, as expected, and that it contrasts sharply with the ratio scale (triangles) determined by asking 12 subjects to estimate directly the apparent duration of the stimulus. The upward curvature of the ratio scale for subjective duration agrees with the results of independent experiments in which the method of fractionation was used to determine the duration that appeared to be half as long as a standard duration (19, 50).

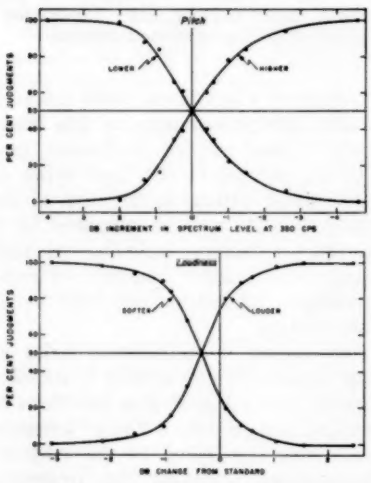


FIG. 2. Psychometric functions from 50 subjects obtained with recorded tests for ability to discriminate the pitch and loudness of bands of noise. Each test contained 110 test items consisting of a standard noise (2 seconds) followed immediately by a comparison noise (2 seconds). The band of noise used in the pitch test was shifted up or down the frequency scale. Note that there is a "time error" for loudness but none for pitch.

The curves in Fig. 1 tell us that on a prothetic continuum the typical subject is unable to equalize the intervals on his category scale, even when instructed to do so. He fails for the basic reason that his ability to tell one magnitude from another varies over the scale and affects the width of his categories. Since he can easily tell 0.5 sec. from 1.0 sec. he tends to put them in different categories; since he can only with difficulty tell 3.5 sec. from 4.0 sec. he tends to put them in the same category. Any procedure that tries directly or indirectly to get the subject to partition a prothetic continuum into equal intervals seems bound to fail in the general case.

For purposes of comparison, the function obtained by counting up j.n.d.'s is plotted in Fig. 1 as a dashed curve. This curve, based on Woodrow's (73) report that the relative j.n.d. for duration is of the order of 10 to 12.5 per cent, shows the form Fechner's law predicts for the scale of apparent duration.

3. *Time-order error.* The so-called "time-error" discovered by Fechner has been pursued in theory and experiment for many decades (23). It refers to the fact that the second of two equal stimuli tends to be judged greater than first. We have reason to believe that a systematic time-order error (as Guilford prefers to call it) is characteristic of judgments on Class I continua. On Class II continua we neither expect it nor do we generally find it.

This hypothesis was first suggested to me by the troubles we met when we

tried to supply the armed forces with recorded tests designed to assess pitch and loudness discrimination for bands of noise (29). Karlin and I discovered that it was easy enough to record a pitch test whose scoring was straightforward, but how were we to score a loudness test when it always showed a time-order error? Were we to try to explain to the military that, although the second stimulus may actually have been 0.1 db lower than the first, if the soldier called it lower they should mark him wrong? We tried to get rid of the error by making the transition from the standard level to the variable level instantaneous, but the error was still there. We finally solved the issue by avoiding it: we dropped the stimulus items whose proper scoring was made ambiguous by the time-order error.

Figure 2 shows results obtained with these tests from groups of 50 subjects each. (In the course of this work several hundred subjects were tested.) The crossing points of the psychometric functions show that there is no time-order error for pitch, but that for loudness the error amounts to about 0.3 db. Similar results were obtained when we used the Seashore pure-tone tests. Postman (43) then took up the problem and confirmed our finding that there is typically a time-order error with loudness but not with pitch.

This perhaps is not the place to try to "explain" the time-order error, but I would like to venture an opinion about it. First of all, it is important to note that the error is typically small—a fraction of a j.n.d.—and hence is close to the limiting "noise level" of our measurements. It is therefore not surprising that the published findings resound with the noise of disagreement. Attempts to map the course of the effect as a function of time are particularly discordant, which is good reason for giving up the name "time-error." Its dependence on time is probably fortuitous. As a matter of fact, in an experiment still in progress (by G. M. Shickman) we have obtained the usual "time-error" when the subject lifts the standard and the comparison weight simultaneously, one with each hand.

What the Fechnerian time-order error really seems to depend on is the same basic process that makes the category scales on Class I (prothetic) continua turn out to be concave downward, namely, the asymmetry of sensitivity—the fact that discrimination is better toward the low than toward the high end of the range. The methods commonly used to measure the time-order error are essentially similar to those we use in category scaling. The subject places stimuli in two or more categories like *heavier, equal,* or *lighter* and the forces that determine his judgments are the same as in any category scaling procedure. (When the method of adjustment is used with loudness a different type of constant error occurs: the variable tends to be set *higher* than the standard [44], but the sign of the error may reverse when the standard is very loud [61].) The relation between the category scale and the conventional time-order error is perhaps best illustrated by means of Fig. 3, where we see three category scales for lifted weights covering different ranges of stimuli. (For a further discussion of these and other such curves, see [64].) The widest range (Cowdrick [7]) gives the usual category curve, concave downward, and we note that the middle weight in the series is assigned a value *above* the middle of the category scale. When the range is shortened (Guilford and Dingman [25]) the same general features are preserved. The shortest range is the one used by Fernberger (11) to determine a difference limen by the method of single stimuli. This curve, like the others, fails to pass

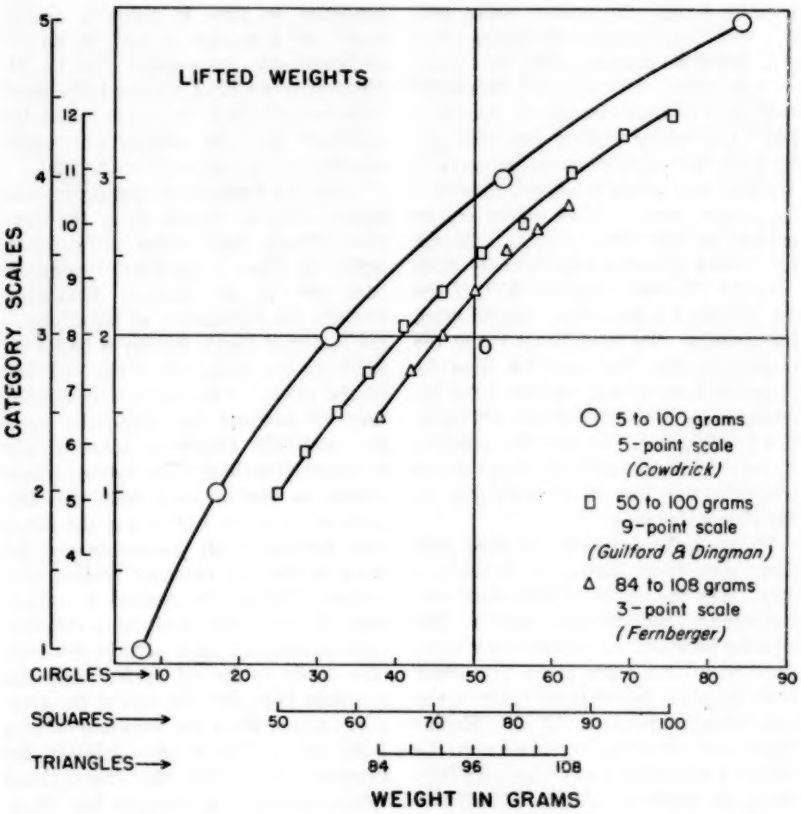


FIG. 3. Category scales for lifted weight. On the middle curve only every other point has been plotted. The end stimuli were designated 4 and 12 but the subject could use other categories if desired. Note that the curves are all concave downward. The fact that they do not pass through the point 0 indicates a "time error."

through the point 0, which is the point defined by the middle of the stimulus range and the middle of the category scale. The distance between the point 0 and the point where the curves cross the horizontal line through the middle of the plot is the time-order error (cf. Pratt [46]). It is as simple as that. By means of the theory of category judgments we can apparently account for Fechner's time-order error without the help of such props as sinking traces, fading images, or assimilation. And it becomes clear why we should expect to find time-order errors on Class I continua (prothetic) but not on Class II (metathetic).

One further point. Not only is the time-order error produced by other factors than time, but it also has nothing to do with order. Nor is it necessarily an error. The name, like the one given to the Holy Roman Empire, is a misnomer in all particulars. If my thought is correct, that we are dealing here with an effect on category judgments that

derives from an asymmetry based on the relativity of discrimination, a better name for it might be the "category effect." Such a label might also help to distinguish this form of bias from other sources of systematic "error," of which there are many. Not all constant deviations in psychophysical experiments can be blamed on the category effect. To take a single example, the observation of Fullerton and Cattell (13) that the second of two equal light flashes appears fainter than the first is probably due to an altered state of retinal adaptation produced by the first flash. In this instance an adaptation effect can apparently override the category effect and produce what would conventionally be called a positive time error. No doubt there are still other sources of systematic bias in experiments of this sort.

It is true of course that time may play a role in these and other systematic errors, in the sense that the effects may vary with the interval between judgments. But our problem is first to try to discover the basic nature of the asymmetries involved in the various kinds of constant errors and then to see how the forces that produce them may vary with time.

4. *Hysteresis.* Since this word means a lagging behind, as when magnetization lags behind the magnetizing current, it seems a good term to describe what happens when the apparent sense-distances between successive stimuli are judged in different orders. The effect shows up in especially dramatic fashion in experiments on bisection and equisection. In a typical experiment on loudness the subject sits before a row of five keys which he presses to produce the tones (1,000 cps). The levels produced by the two end keys are fixed, 40 db apart, and the subject adjusts the levels controlled by the intermediate keys in order to divide a 40-db interval into four equal-appearing steps in loudness.

Where the subject sets the levels depends on whether he listens to the loudnesses in ascending or descending order. I found that in the ascending order he sets the bisecting level some 5 to 8 db higher than in the descending order. In an experiment run to verify this effect, Garner (15) found an average discrepancy of 5.8 db. It is as though the loudness the subject hears lags behind what he should hear as he goes up or down the scale. As a result of this lag the graphs of the functions shown in Fig. 4 exhibit a "hysteresis loop": the ascending path is different from the descending path. Please note, however, that in calling this phenomenon hysteresis I am trying to describe it, not explain it. I am not sure I know how to explain it.

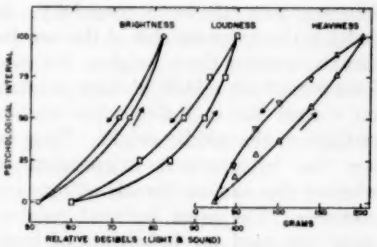


FIG. 4. Hysteresis effects in judgments of intervals on three sensory continua. The arrows indicate the order of stimulus presentation. For brightness and loudness the subject adjusted the middle stimuli to divide the range into two or four equal-appearing intervals. For heaviness the subject lifted three weights and indicated the relative apparent position of the middle weight by means of a slider between two markers on a bar. Five different middle weights were used in ascending and descending order.

As shown in Fig. 4 the same hysteretic effect occurs in the bisection of brightness intervals and in judgments of the intervals between lifted weights. Partly because the range was shorter (30 db) the brightness bisections show less hysteresis than loudness. Lifted weights, on the other hand, show relatively more. The possibility suggests itself that the magnitude of the hysteresis may vary

inversely with differential sensitivity. In judging the apparent size of the two intervals between three weights, the subjects moved the middle of three pointers on a steel bar to indicate the relative position of the middle weight. Thus we see that hysteresis is independent of whether the subject himself adjusts the stimulus. The same bar and pointers were also used for judgments of loudness, and the resulting hysteresis was essentially similar to that shown in Fig. 4. The data in Fig. 4 are typical examples from an extensive series of unpublished experiments, some of which were run by R. J. Herrnstein.

Hysteresis occurs on the three examples of Class I continua that we have tested. The hypothesis I am suggesting is that it may occur on all continua of this class, but that it probably does not occur on continua of Class II. Evidence for this possibility is suggestive, but not yet conclusive.

Bisections on the continuum of visual position, for example, seem to be free of hysteresis. At least what minor effect has been found is in the opposite direction from that observed on Class I continua. Volkmann kindly undertook to test this matter on the large viewing screen (21.6 × 6.8 ft.) at Mt. Holyoke College (6). The subject sat 35 feet from the screen and adjusted the position of the middle of three points of light that came on in the sequences right-middle-left and left-middle-right. The end lights were a meter apart. When the 12 observers were instructed to fixate the center of the screen, the two orders produced an average difference in the bisection point of only 0.25 cm. When the instructions were to look at each stimulus as it came on, the discrepancy was 3.6 cm. Since both discrepancies were in the opposite direction from those observed with loudness, brightness, and lifted weights, we may conclude that, whatever constant errors may occur on this Class II continuum, they apparently constitute a different phenomenon from the Class I hysteresis.

The question of hysteresis in pitch judgments is less easy to answer. We found no evidence for hysteresis in the experiment that led to the construction of the mel scale (67), but the fact is we were not looking for it. In a more direct test of the issue I later had 10 subjects bisect various pitch intervals in ascending and descending order. The experiment was casual and not definitive, and the results were not clear cut. For some subjects hysteresis seemed to be present, for others not. Two subjects happened to have "absolute pitch"—the best cases I happen to have encountered. Each of these two subjects set the middle tone to the same frequency on every trial, regardless of the order of listening. So it seems that at least some people show no hysteresis in their pitch judgments. But this is apparently not true of others.

In a recent laboratory exercise, two of my students, D. D. Greenwood and M. L. Israel, found hysteresis in the equisections of all but one of nine subjects who divided the interval 400–7,000 cps into four equal-appearing intervals.

Incidental evidence for the absence of hysteresis in pitch bisections is found in an experiment by Cohen, Hansel, and Sylvester (5). They were concerned with another problem, but part of their procedure called for the bisection of the pitch distances between 1,000 and 3,000 cps and between 2,000 and 4,000 cps in both ascending and descending order. The bisections were as follows:

For 1,000–3,000 cps, going up: 1,874 cps; going down: 1,838 cps.

For 2,000–4,000 cps, going up: 2,693 cps; going down: 2,808 cps.

The average standard deviation for all settings was 303 cps.

We see that the interval 1,000–3,000 cps shows an insignificant discrepancy (36 cps) and that the larger discrepancy

(115 cps) shown by the interval 2,000–4,000 cps is less than half the standard deviation. It is also important to note that the directions of the discrepancies between the results for ascending and descending orders are opposite in the two parts of the experiment. Incidentally, the points of bisection that are predicted by the mel scale are 1,859 cps and 2,280 cps, which are within 3 per cent of the observed values.

All in all then, the evidence for hysteresis in pitch bisections is ambiguous. The pitch continuum probably means different things to different people, especially to those with and without "absolute pitch." It is in many ways a difficult continuum about which to generalize (see, for example [64]).

Another point concerning bisection deserves mention. On continua of Class I (prothetic) the point of bisection falls consistently below the point predicted by the ratio scale of the subjective magnitude (39, 58, 63). This is true of the combined results from which hysteresis has been eliminated by an averaging of the ascending and descending judgments. It is also clearly evident in the early experiments by Fullerton and Cattell (13) who explored the "force of movement" by the methods of halving, doubling, and bisection. This error in bisection is scarcely surprising when we consider that bisection and category rating-scales have much in common. In both instances the subject tries to equalize intervals, and in both instances he misses the mark in the same direction—presumably because both kinds of judgments are controlled partly by discrimination, and discrimination varies from one end of the continuum to the other.

On continua of Class II (metathetic), on the other hand, the results of bisection agree with those obtained by fractionation and by magnitude estimation. This has been demonstrated for pitch (64, 67) as well as for inclination or "bearing" (47, 64). Apparently it is only on Class II continua that the construction of a scale of subjective magnitude by the method of bisection leads to a valid outcome, for only there can bisection be shown to agree with direct magnitude estimation.

In our experiments on weight and loudness in which the subject moved a pointer along a bar, the parallel with category rating-scales is even more obvious. In effect, the bar and pointer constitute a continuous rating scale, analogous to the widely used scales on which the rater marks a point on a line with a pencil. The bar and pointer also give results (hysteresis averaged out) that differ from those predicted by the ratio scale of the Class I continuum.

We see then that possibly four functional criteria can be used to divide the perceptual continua into two general classes. Clear examples of each of these classes are easy to find, but this is not to say that all examples one might think of will prove easy to classify. When we try to impose a simple order on the complex fabric of our experience there is an ever-present risk that we will oversimplify what is not simple. Class I (prothetic) constitutes a rather unitary and well behaved group of continua, but Class II (metathetic) includes various disparate aspects of things, many of which do not fall naturally onto unitary monotonic scales, and some of which may not even be orderable. Some continua may turn out to be hybrid mixtures of both Class I and II. Pitch, in fact, may be one of these (64).

In most of what follows we will limit our concern to the prothetic continua (Class I), the "quantitative" aspects of things, for it is here that we can most profitably examine the principles that psychophysics has set itself to devise. The first of these principles concerns the relation between stimulus and sensation.

THE POWER FUNCTION

There is a growing bundle of evidence to indicate that on prothetic continua the form of the "psychophysical law" is a power function. Most of this evidence, accumulated since 1930, is the product of a reoriented psychophysics that brings "direct" methods to bear on the Fechnerian query regarding the relation between stimulus and sensation. These direct methods lead to ratio scales of perceptual magnitude, and to a first approximation these scales show that the psychological magnitude is a power function of the stimulus magnitude.

The basic principle seems to be that equal stimulus ratios tend to produce equal sensation ratios. Empirical evidence to support this principle has been obtained for at least a dozen perceptual continua, with the aid of several different experimental methods. What the principle entails can be illustrated by examples such as these: In order to make one sound seem half as loud as another, the physical energy must be reduced by about 90 per cent (10 db), and this required reduction is approximately the same regardless of the level from which we start. The same is true of brightness, and the required reduction is of the same order of magnitude as for loudness. In order to make one lifted weight seem half as heavy as another, the original weight must be reduced by about 38 per cent, and this percentage reduction is approximately constant over a wide range of stimuli. The approximate constancy of the percentage reduction corresponding to a given subjective ratio demonstrates that the sensation ψ is proportional to the stimulus S raised to a power n. Thus

$$\psi = kS^n.$$

When we convert this equation to logarithmic form we obtain a linear equation which has a certain practical usefulness, for the function can then be represented by a straight line in log-log coordinates. The slope of this line corresponds to the exponent n.

When we know the value of ratio r between the stimulus judged half and the standard stimulus we can obtain the value of n by dividing log r into log 0.5. In general, if s is the sensation ratio (determined by the method of ratio production) and r is the corresponding stimulus ratio, then

$$n = \frac{\log s}{\log r}.$$

It is not to be expected, of course, that a simple power function will hold from zero to infinity. What the limits may be must be decided by experiment. We know that some functions relating stimulus to response exhibit discontinuities (e.g., apparent numerousness and flash rate), and we know that our formulas might need eventually to take account of second-order perturbations in the vicinity of the absolute threshold (cf. **8, 58**). But our present concern is with first-order approximations to general laws.

METHODS FOR RATIO SCALING

Before we consider the concrete instances of ratio scales, let us look briefly at the problem of procedure. Methods for constructing ratio scales of subjective magnitude are a relatively new development. It is true that Merkel (**31**) with his *Methode der doppelten Reize* (1888) tried to find the stimulus that appeared to double the sensation, but Merkel's effort seems to have had little influence on psychophysics. Fullerton and Cattell (**13**), who used the methods of doubling and halving, fared little better. Even Titchener (**72**) in his thorough and scholarly Instructor's Manual makes only off-hand allusion to Merkel's method of "doubled stimuli." And of the author of the method he says "we are reminded that he is, by mental constitution, a physicist rather than a psychologist . . . and that his work as a

whole is exceedingly poor in introspective data" (p. 223). It may be that "mental constitution" plays a role in these things, as Titchener suggests. I wonder what he would make of the fact that at least seven different physical laboratories have made important contributions to the development of the ratio scale of subjective loudness—the sone scale—as against perhaps three different psychological laboratories.

The impetus for the development of a method is a problem. Method for its own sake, tempting and fascinating as it may be, often leads to little—except methodology. Physicists and psychologists went to work to refine the procedures for measuring loudness mainly because the substantive outcome was of interest and importance to someone, particularly to acoustical engineers. This is shown by the fact that some of the earliest studies were paid for by commercial companies. Curiously enough the practical problem originated in an obvious failure of Fechner's law. Not long after they had adopted the decibel scale for measuring sound intensities, the engineers noted that equal steps on the logarithmic (decibel) scale do not behave like equal steps, for a level 50 db above threshold does not sound at all like half of 100 db, as Fechner's law implies it should. Since the acoustical engineer must often interpret to his customers the meaning of esoteric acoustical measurements, it soon became clear that a scale was needed on which the numbers would be proportional to how loud things sound to the typical listener. Without the motive of this kind of practical problem to spark the development of method, it is a fair guess that the ratio scales for the fourteen perceptual continua we will examine might never have been constructed. How a subjective ratio scale, once achieved, can be put to work is shown by the way we were able to use the sone scale in the development of a procedure for calculating the loudness of a complex noise from a spectral analysis of the sound (61).

The methods for constructing ratio scales are still in the process of development, but in one form or another they all require subjects to make quantitative estimates of subjective events. Many authors have screamed that this is nonsensical, meaningless, and impossible (12, 34), but those who follow these methods go ahead and do it anyhow. These direct assessments of sensation seem not so impossible after they have been made.

At present the methods are principally four, but each has subvarieties. We may classify the methods more or less systematically as follows:

1. Ratio estimation
   a. Direct judgment of ratios
   b. "Constant-sum"
2. Ratio production
   a. Fractionation
   b. Multiplication
3. Magnitude estimation
   a. Prescribed modulus
   b. No designated modulus
4. Magnitude production

I am well aware that few activities in psychology exhibit greater terminological confusion than the naming of methods, and that there is probably no effective cure for the burgeoning, ramifying process that occasions the welter of names. The best we can do is take stock from time to time and attempt a little systemization. Having formerly contributed my full share to the confusion, in the foregoing list I have tried for a dose of order.

Partly through my efforts the second class of procedures, which was the first used historically, got the name fractionation (68), because the usual procedure has been to require a subject to set one stimulus to produce a sensation half as great as a given standard. Frac-

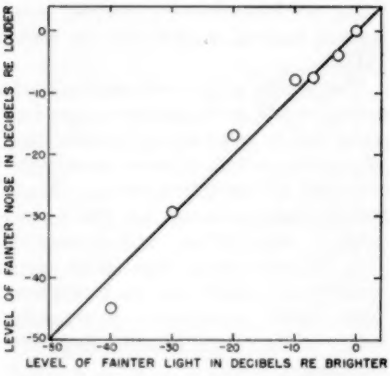


FIG. 5. Results of adjusting a loudness ratio to equal an apparent brightness ratio. The subject viewed a pair of luminous circles of which one was dimmed by an amount shown on the abscissa. He also pressed a pair of keys to produce two different levels of white noise. He adjusted one level (ordinate) to make the loudness ratio seem equal to the brightness ratio. The brighter light was about 99 db re 10^{-10} lambert and the louder noise was about 92 db re 0.0002 microbar.

tions other than a half are also used with consistent results. But fractionation in this sense is only one facet of a more general procedure. The other facet, which we might call *multiplication,* involves the complementary process of requiring the subject to identify or produce a prescribed ratio that is greater than one, e.g., to double, triple, etc., a given standard. This procedure has not been used very much, probably not as much as it should be, for there is abundant evidence (62) to indicate that, when it is made to complement halving, the procedure of doubling can help to balance out certain systematic biases.

The two procedures together may be called *ratio production.* The production of ratios can be carried out by a variety of different techniques. Thus the experimenter may allow the subject to adjust a stimulus to produce a prescribed ratio to a standard, or the experimenter may set the stimulus and ask the subject whether it meets the prescribed ratio (method of "constant stimuli").

An interesting variant of ratio production involves the fixing of two luminances to define an apparent ratio which the subject is to reproduce by setting two loudnesses to the same apparent ratio. In an experiment by J. C. Stevens in this laboratory the subjects set nearly the same *physical* ratio (decibels) between the intensities of the white noises as the experimenter set between the intensities of the two white surfaces. Figure 5 shows the median settings made by 15 subjects, each of whom twice adjusted the loudness of the second of two noises until the apparent ratio between the noises equaled the apparent ratio between two luminous targets viewed against a dark surround. The subject sat about 4 feet from the targets, which were 4.4 cm. in diameter and separated by about 17.5 cm. The fact that the median settings of the variable noise fall fairly close to the line indicates that the subjective impressions produced by white light and white noise are similar functions of stimulus intensity. The adjustments made to the largest ratio (40 db) were of course quite variable: the semi-interquartile range was about 10 db.[2]

The method of *ratio estimation* is the inverse of ratio production. Instead of prescribing the ratio in advance, the experimenter presents two (or more) stimuli and asks the subject to name the ratio between them. The subject may name the ratio directly, as in the pioneer experiment of Richardson and Ross (49), or he may be constrained to express the ratio by dividing a given number of points between the two stimuli in the manner proposed by Metfessel (32). The constraint involved in the so-called "constant-sum" method has

[2] For the rationale of measuring luminance in decibels see (59).

obvious disadvantages when large ratios are involved.

The method of *magnitude estimation* (**60**) dispenses with ratios as such and requires the subject to assign numbers to a series of stimuli under the instruction to make the numbers proportional to the apparent magnitudes of the sensations produced. The experimenter may prescribe a modulus by presenting a given stimulus and calling it some particular value, e.g., 10, or he may leave the subject free to pick his own modulus. (Note: if the magnitude estimations give a skewed distribution, as they usually do, it is advisable to calculate medians instead of means.)

The method of magnitude estimation has a logical inverse in *magnitude production*—a possible method that has thus far been mostly ignored. Instead of presenting a series of stimuli in irregular order and asking the subject to judge their apparent magnitudes, the experimenter might name various magnitudes and ask the subject to adjust stimuli to produce proportionate subjective values. Like any given procedure, this method probably has its "bugs" and biases and it might be interesting to find out what some of them amount to.

One thing we know is that in using this method the experimenter must resist whatever impulse he may have to fix or designate the top and bottom of the range, for this will change the problem into one involving category scaling. In some earlier experiments (**64**) we used this procedure, which we may call category production, to generate a 7-point category scale of loudness. We presented two levels which we designated 1 and 7 and then asked subjects to produce each of the other categories in irregular order. The results are like the usual category judgments obtained on Class I continua: the function is concave downward when plotted against the sone scale.

Magnitude production of a sort has been used in an experiment in which we used a brightness rather than a number to designate a magnitude. We set the luminance of a single target to various levels and asked subjects to adjust a noise to make its loudness appear as great as the brightness of the light. Although this investigation is still in progress (by J. C. Stevens) the results appear to be fairly consistent with what we know about the subjective scales for loudness and brightness. The intensities of the white noise are made roughly proportional to the intensities of the white light, again suggesting that loudness and brightness are similar functions of intensity.

All four of the methods listed above provide the kind of data necessary for the construction of a ratio scale (**55**). Each method can of course be varied and modified in numerous ways. Not only do the methods need to be altered and adapted to fit the problem at hand, but in any serious effort to establish a definitive scale for a given perceptual continuum it is desirable to ferret out the possible sources of bias by using different methods and by altering relevant parameters. In the present state of the art a valid scale that is representative of the typical subject can scarcely be hoped for from a single try.

The following ratio scales of subjective magnitude have been erected by means of one or more of the general procedures named above. They have not all been given the intensive study and cross-checking they deserve, but our interest here is in their general form rather than in their detailed adequacy. To a first approximation they are all power functions.

FOURTEEN RATIO SCALES

Our review of the fourteen examples of ratio scales of subjective magnitude will be brief, because most of the details

TABLE 1

PERCEPTUAL CONTINUA ON WHICH PSYCHOLOGICAL MAGNITUDE IS A POWER FUNCTION OF THE STIMULUS

The second column shows the approximate exponent of the power function. Names suggested for the various subjective units are listed in Column 3. The methods used (Column 4) are indicated by numbers that refer to the methods listed elsewhere in this paper.

Continuum	Exponent	Name of Unit	Methods Used	Reference Source
Loudness	0.3	sone	1a, 2a, 2b, 3a, 3b, 4	(58)
Brightness	0.3–0.5	bril	2a, 2b, 3a, 3b	(28), (64)
Visual distance	0.67		2a	(17)
Taste	1.0	gust	2a	(1)
Visual length	1.1	mak	1a, 2a, 3a	(8), (47), (64)
Visual area	0.9–1.15	var	1a, 1b, 2a	(8), (47)
Duration	1.05–1.2	chron	2a, 3a	(19), (50), 64)
Lightness of grays	1.2		3a, 3b	(64)
Finger span	1.3		3a, 3b	*
Numerousness	1.34	numer	2a	(69)
Heaviness	1.45	veg	1a, 1b, 2a, 2b, 3a, 3b	(64)
Auditory flutter	1.7	flut	2a, 2b	(42)
Visual velocity	1.77		2a	(9)
Visual flash rate	2.0		2a	(48)

* S. S. Stevens and Geraldine Stone. Research in progress.

are available elsewhere. We can help the cause of brevity by means of Table 1 where the significant facts are listed and the sources of more detailed information are tabulated. In Table 1 the continua are ordered by increasing size of the exponent of the power function, and whenever a name for the subjective unit of the ratio scale has been suggested it is recorded. The methods that have been applied to each scale are indicated by numbers that refer to the listing in the previous section. I think all these continua belong to Class I, but this of course may be open to question.

Since Table 1 was prepared, an experiment on smell has been started in this laboratory by T. S. Reese. The early results suggest that the apparent intensity of the odor of benzaldehyde (synthetic almond) is a power function of the concentration in moles per liter of air. The exponent seems to be of the order of 0.2.

It goes without saying that all these scales have not been determined with equal care and thoroughness. About some of them we know a lot and have reason for confidence in their form; others have received only limited attention. Some of these attributes are easy to scale; others are more resistant to precision. Some of them represent perceptions about which we frequently make judgments of a more or less quantitative sort and about which the subjects in an experiment know, or think they know, so much about the stimulus that the "stimulus error" (3) may be a ready danger; others involve stimuli whose physical parameters are so unfamiliar to the typical subject that his temptation to try to estimate a physical measure of the stimulus, rather than to judge how it appears, is remote and uncompelling. Yet all these ratio scales approximate power functions.

I say approximate because it is certain that not all judgmental continua are strictly power functions of an arbitrary physical parameter of the stimulus. We know, for example, that over

most of the range the loudnesses of a 1,000-cycle tone and a white noise grow with intensity according to similar laws, but at very low intensities the loudness of the white noise grows more rapidly. Hence the loudnesses of noise and tone cannot both follow power functions throughout their entire range, and it is even doubtful that either is strictly a power function at better than a first-order approximation (62).

On this question of the form of a psychophysical law we ought perhaps to try to learn a lesson from Fechner's mistake and forego the temptation to insist on the exactitude of a particular formula. In the realm of human judgment we may aspire to the discovery of first-order invariances, but it would be idle to pretend that precision can be pushed to the level envisaged by Fechner. The judgment of subjective magnitude is inherently a "noisy" phenomenon. When people try to describe a sensation in quantitative terms they face a difficult task, and the factors that affect the outcome are numerous and subtle. Patience and experimental skill can probably clean up part of the variance, but there will always remain irreducible dispersions to set a level below which we sink into uncertainty. In the broader perspective of things the power law stands out as a first-order relation, just as Weber's law describes the first-order relation between magnitude and resolving power. Although we are sometimes more fascinated by second-order departures from first-order relations than we are by first-order relations themselves, the first-order generalities have the broader significance. It is of greater moment to know the first-order law for the velocity of falling bodies than it is to prove that a host of variables can produce departures from the law.

Brief comments on some of the scales listed in Table 1 are now in order. We have already discussed loudness in various connections, and a review of most of the work on the problem is available (58). An example of a segment of the loudness function obtained from 32 subjects by the method of magnitude estimation is shown in Fig. 6. It is interesting that the sone scale is, so far as I know, the only subjective ratio scale for which the American Standards Association has appointed a committee to recommend a standard form.

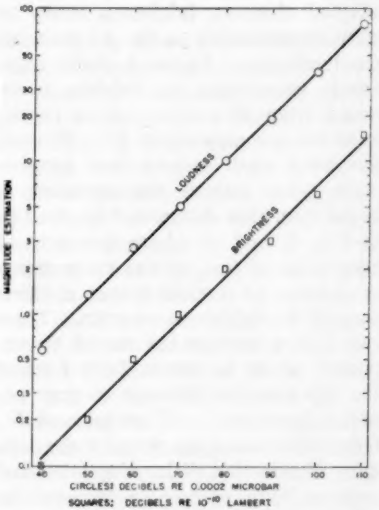


FIG. 6. Median magnitude estimations for loudness and brightness. For loudness each of 32 subjects made two estimates of each level (1,000 cps) in irregular order with no designated standard. Estimates were transformed to a common modulus at the 80-db level. For brightness each of 28 dark-adapted subjects made two estimates of each level in irregular order. The target subtended about 5° and was illuminated for about 3 sec. at each presentation. Once at the beginning of each session the subject was shown the level 70 db (14 subjects) or 80 db (14 subjects) and told to call that level 10. Estimates for all subjects were transformed to a common modulus at 70 db.

Brightness is still under active study. For ordinary luminous targets seen against a dark surround by a dark-

adapted observer, brightness seems to grow approximately as the 0.3 power of the luminance. Figure 6 shows magnitude estimations for brightness obtained from 28 subjects in an experiment being conducted by J. C. Stevens. For very small targets that approximate point sources the exponent is larger than that determined by the line in Fig. 6, and of course the state of adaptation of the eye and the presence or absence of contrast makes a difference to the brightness function. These two factors increase the size of the exponent, as can be seen in Table 1 where the exponent for lightness of gray papers is listed as 1.2. When grays of different reflectances are viewed under ordinary illumination, light adaptation and contrast both operate to increase the steepness of the brightness function to a different order of magnitude from that observed with luminous targets seen in the dark.

Visual distance refers here to the apparent distance from the observer to a marker on a 200-foot line that stretched away from the observer on a large flat lawn. Two subjects made fractionation judgments. The combined results can be fitted by a $\frac{2}{3}$ power law as well as, if not better than, by the formula proposed by Gilinsky (17). Specifically, the data in her Fig. 6 fall close to a straight line. The main difficulty with the fit occurs at the shorter distances, where for one observer the apparent half values are nearly equal to the objective half values. But we would expect apparent distance to approximate objective distance when the angle between the line on the grass and the observer's line of regard becomes large. Perhaps if the line went out from the bridge of the observer's nose we would obtain a power function all the way—with an exponent less than one. Of course apparent distance, like apparent size, is specific to the experimental situation and is affected by cues of various sorts. It is not to be expected, therefore, that a single power function will govern all judgments of distance. As a matter of fact, in a very different set-up involving short distances (4.5 meters and less) Gruber (21) obtained data that fit a power function with an exponent of approximately 1.3.

Taste, curiously enough, gives an exponent of approximately unity when the stimulus measure is the ratio of solute to solvent (1). The functions reported for four different taste substances vary somewhat from one to another, for taste is a troublesome modality to work with, but they all give fairly straight lines in a log-log plot and the slopes hover around the value one. Another experiment (on sweetness [30]) gave only a rather poor approximation to a power function.

Length and duration give subjective ratio scales that are nearly proportional to the physical scales on which we usually measure the stimulus. On balance, however, there is a tendency in both cases for the exponent of the power function to be greater than one, which means that on these continua the magnitude of the subjective impression is a slightly accelerating function of the stimulus magnitude. The accelerating function for duration is shown in Fig. 1.

Apparent area seems to be nearly proportional to physical area, but the shape and size of the area to be judged seem to produce second-order differences. For smaller squares the exponent appears to be greater than for larger squares. In a recent experiment with circles the exponent turned out to be less than one (8).

For continua such as length it is sometimes held that learning plays a major role: we learn what length measures half as long as another, and so it comes to look half as long—which is one way of explaining why the exponent of

the power function is near unity. This may or may not be true. Of course, if by learning we mean that the organism has had a history, we can never hope to experiment on subjects who have not learned. How the organism got to be the way it is will always be an engaging problem, but it is also a worthy object of attention to discover just how the typical subject now behaves, regardless of how he got that way.

The "learning explanation" of these ratio scales seems much less convincing for such continua as loudness and brightness. Despite Ebbinghaus' argument (see [60]) that we learn the meaning of a doubled brightness by illuminating the scene with two lights instead of one, the typical viewer actually requires some nine or ten times the illumination before he reports the brightness doubled. Why is this? And why is it that when the reflectance of gray papers is involved, where it is scarcely conceivable that the typical viewer knows the nature of the physical processes at work in the stimulus, the exponent is greater instead of much less than unity? Sophistication as regards the stimulus seems less than a likely explanation.

It is true, however, that one can learn to estimate the physical magnitude of a stimulus, as a good photographer well knows. He corrects what he perceives by what he knows about light and its relation to subjective impressions. By a similar procedure the acoustical engineer may estimate decibel levels with fair precision. He learns what a given subjective impression means in terms of the indications on a level meter. I have tested at least a half dozen acoustical engineers whose business involves the estimation and measurement of decibel levels. Despite their familiarity with decibels their subjective scales of loudness are not systematically different from those of other listeners. They can tell me, for example, both that one sound seems about half as loud as another, and also that the one is about 90 db and the other about 100 db.

Let us return now to Table 1. The ratio scale for finger span was obtained as part of a program that is still in progress. Subjects judged the apparent thickness of blocks of wood placed for about a second between the thumb and middle finger. The magnitude estimations obtained show that over a range of from about 0.2 to 6.0 cm the impression of felt thickness grows as a power function of the thickness of the blocks, and that the exponent is about 1.3.

Numerousness refers to the impression one gets from looking at a collection of dots that contains too many to be counted. The exponent of about 1.34 was derived from Taves' experiment in which he used the method of fractionation. I computed the exponent from the mean fraction judged half for all standards above 8 dots. The exponent is equal to log 0.5 divided by the logarithm of the fraction judged half.

Lifted weights have been scaled by several methods in several laboratories and most experimenters have found close approximations to power functions. The value of 1.45 for the exponent of the veg scale was obtained from pooled results from five different laboratories. For the procedures involved, see (64). It is interesting to note that when Fullerton and Cattell in 1892 had their subjects try to halve and double the apparent force exerted on a dynamometer they came quite close to the values that would be predicted by the veg scale. The two values obtained fell on opposite sides of the predicted value, and the largest discrepancy was about 10 per cent of the predicted value. Although these data are meager they suggest that the apparent force of this type of muscular pull may follow a power function that is like the veg scale.

Auditory flutter refers to the appar-

ent rate at which a white noise is turned on and off. Over a wide range of flutter frequencies (5 to 200 per second) the apparent rate of flutter grows as the 1.7 power of the frequency of stimulus interruption. Pollack used both fractionation and multiplication in determining this function. The one value he tested at a rate below 5 per second suggests that there may be a discontinuity in the function at about this point. Otherwise the data give a close fit to a power function.

Velocity of seen movement gave data that fit a power function so well that, as the authors say, "The goodness of fit may be judged as excellent." The subjects alternately viewed two horizontally moving targets consisting of vertical white and black bands, and they adjusted the velocity of one to make it appear half the other. The standard velocities ranged from 1.8 to 42 mm/sec.

Ekman and Dahlbäck (9) call attention to another interesting fact, namely, that when the subjects were asked to set one velocity *equal* to a standard velocity the variability of the settings was not very different from the variabilities obtained when one velocity was set to appear *half* of a standard. They say, "the degree of uncertainty in fractionation is only a little greater than that in equating stimuli." Hanes (28) had earlier noted the same thing in his experiments on brightness. He says, "Extremely interesting is the fact that the standard deviations around the group means of the equality matches . . . are of the same order of magnitude as the deviations around the means of the $\frac{1}{2}$ and $\frac{1}{3}$ estimates at the same brightness levels." I can add for loudness that the variability of halvings and doublings is actually slightly less, in general, than the variability of equality matches made between two noises of substantially different spectra, say, a hiss and a rumble. In experiments involving several thousand loudness matches (61) the standard deviations ranged from 1.5 db, for very similar spectra, to 9.0 db for disparate spectra. Comparable variabilities for 36 subjects who halved and doubled the loudness of a white noise averaged 4.2 db (45). For pure tones, Garner (14) reports variabilities for halving that range from about 3 to 6 db.

Visual flash rate gives the largest exponent yet encountered. For rates between 5 and 10 per second the exponent is about 2.0. Actually the subjective scale for flash rate seems to break up into two segments: one for slow rates where the subject probably judges time elapsed between flashes, and one for faster rates (greater than about 5 per second) where the subject seems to judge rapidity of flash. Except near the transition point, both segments of the scale approximate power functions with slopes decidedly greater than unity.

This completes the list. Let us now look briefly at some of the antecedents of our power law.

SOME ANTECEDENTS

The idea that the relation between stimulus and perceptual response might be a power function is no new thing. What is new is the devising of experiments to demonstrate the fact.

The first recorded notion that a power function might be involved appears to be the conclusion reached by Plateau (41) in the 1850's—before the appearance of Fechner's *Elemente* (1860). Plateau reasoned that, since the apparent relations among different shades of gray remain *sensiblement le même* when the general illumination is changed, the *ratios* among the sensations produced by the grays must remain fixed. This, he said, and quite rightly, is more rational than Fechner's view (later revived by Wright [75]) that it is the *differences* that remain constant. Plateau's assumption entails a power function,

Fechner's a logarithmic function. By the present evidence, Plateau was right and Fechner was wrong. Sad to relate, however, Plateau later changed his mind, and for a reason we now can see was irrelevant to the issue: Delboeuf's equisection experiment did not yield a power function. And Fechner, crafty polemicist that he was, disposed of Plateau's power function by the equally irrelevant argument that its champion had repented.

But the power function came back to plague Fechner by another route. Brentano (4) suggested that an increase in a sensation is just noticeable when the increase is a fixed proportion of the original sensation. Brentano did not explore the mathematical consequences of this introduction of a kind of "Weber's law" into the realm of sensation, but Fechner did the mathematics for him (10) and showed that, if Weber's law holds for *both* the sensations and the stimuli, the consequence is a power function. This way of getting to a power function was later argued for by others, particularly Grotenfelt (20), but since there is scarcely any way to prove the basic assumptions involved, the cause was lost under Fechner's massive attacks upon it.

A third approach to the power function was devised by Guilford (22, 23), who called it the "nth-power law." It is based on the Fechnerian premise that j.n.d.'s are subjectively equal and that the scale of sensation can be had by summing them up. Guilford's new twist was to alter Weber's law by making the stimulus increment $\triangle S$ proportional not to S but to a power of S. A Fechnerian integration then yields a power function—which demonstrates that we can sometimes reach a correct conclusion by starting from two wrong assumptions. The assumption that $\triangle I$ is a power function of I, with an exponent lying between 0.5 and 1.0, is not very easily justified by the facts. It is not so easily justified as the more plausible assumption that $\triangle S = k\ (S + d)$ where k and d are constants. This latter formula fits the data Guilford himself cited better than does the power function, and there is other evidence to support it (33).

But it is Guilford's other assumption that is the greater source of trouble. How serious is the consequence of assuming the equality of j.n.d.'s becomes clear if we begin with any of the power functions (Table 1) whose exponent is greater than 1.0 and follow Guilford's development in the reverse direction. We reach the startling conclusion "that $\triangle S$ would be a decreasing function of S, which is unheard of in psychophysical research" (24).

In the light of what we now know, it would appear that of the three historical approaches to the power function Plateau's was based on the most valid assumption. But unfortunately Plateau's premise was only an assumption, and its empirical verification had to wait for nearly a hundred years.

Our position now is more fortunate. The road to the formulation of a general psychophysical law need no longer be strewn with assumptions about j.n.d.'s or contrast steps. The principle I am proposing, that equal sensation ratios are produced by equal stimulus ratios, is merely the summary statement of what we observe on at least a dozen perceptual continua. The power function which this simple principle entails appears to be the first-order relation between stimulus and response. Second-order departures from this law are certain to exist, and their exploration may sometimes prove interesting and instructive, but the wide invariance of the first-order relation deserves to remain the matter of first importance.

THE SUBJECTIVE SIZE OF THE J.N.D.

The issue raised by Fechner when he assumed that each j.n.d. represents a constant increment in sensation is a stubborn and vexatious problem. It is equivalent to the assumption that measures of resolving power provide equal units that can be used as a measure of magnitude. Applied to Class I (prothetic) continua the assumption is dead wrong, but how can we prove it? Or rather, how can we make the proof persuasive and convincing? Proof has been offered in the past (**33**, **38**, **52**) but ingenious counter-arguments have been marshaled against the evidence. Piéron (**40**) is the objector most persistent and erudite, but others join the chorus. It seems so simple and sensible to take the j.n.d. as a unit of sensation. How can it not be proper?

It is improper simply because it is wrong. Let us consider a concrete case. Both fractionation and magnitude estimation show what is quite obvious to begin with, namely, that the apparent length of a line seen against a homogeneous background is very nearly a linear function of physical length. Two inches look about twice as long as one inch. So we know that subjective length is related to physical length by a power function whose exponent is close to unity. Now what about the j.n.d., the resolving power? We can easily tell visually when one centimeter is added to one centimeter, but we can scarcely tell when a centimeter is added to a meter. So resolving power must be relative, or approximately so. As Münsterberg (**37**) showed, Weber's law holds fairly well for length. Now we put these two facts together and what happens? We see immediately that in terms of their psychological magnitude, as measured by the scale we first set up by fractionation and magnitude estimation, the j.n.d.'s get larger as we go up the scale. Q.E.D., the j.n.d. is not constant in subjective size, and counting them off does not yield the magnitude scale.

The principle involved can be clearly seen if we put the problem in more general terms. Let us assume that sensation ψ is a power function of stimulus magnitude S and also that Weber's law is true, i.e., $\Delta S = kS$. By the first assumption we may write, disregarding constants that depend only on the choice of units,

$$\psi = S^n.$$

If Weber's law is true, when we count off the j.n.d.'s and record their number J as a function of the stimulus, we obtain

$$J = \log_c S,$$

where c is a constant that can serve as the base of the logarithm. We can rewrite this equation as

$$S = c^J$$

and substitute this expression for S in the first equation, which gives

$$\psi = c^{Jn}.$$

This equation tells us that the psychological magnitude grows as an exponential function of the number of j.n.d.'s J, which brings us to the interesting conclusion that the scale of sensation is an exponential function of the scale Fechner believed it to be. If we differentiate this equation we get

$$\frac{d\psi}{dJ} = ac^{Jn},$$

where a is a constant. The equation tells us that instead of being constant as Fechner assumed, the subjective size of the j.n.d. grows as an exponential function of the number of j.n.d.'s above threshold. Simple substitution in this last equation also reveals that the sub-

jective size of the j.n.d. is proportional to the psychological magnitude ψ.

These equations tell us how it would be if Weber's law were strictly true, but since this law is not strictly true the situation is not quite so simple. The complications, however, are only second order. A superior form of Weber's law is $\Delta S = k(S + d)$, which was known to Fechner, Helmholtz and others. By the addition of the small constant d, this form makes allowance for the inevitable residual "noise level" in the perceptual system, but at the same time it preserves the basic principle that resolving power is relative. If this form is accepted, the foregoing equations and derivations get complicated, but the conclusions they lead to are not substantially altered. It is still true that the subjective size of the j.n.d. grows rapidly as we go up the scale.

Actually, since the value of the constant d is relatively small, its addition to the formula for Weber's law affects the functions in a significant way only at low stimulus levels. At levels near threshold the general formulas developed above break down, but over the middle and higher stimulus ranges they give a good account of things.

A concrete instance of these general relations is shown graphically in Fig. 7. Miller (33) measured the j.n.d. for white noise by the so-called quantal procedure (65) which provides an estimate of the quantal increment in the stimulus that is just detectable. The results are well described by the relation $\Delta I = k(I + d)$. Miller summated the quanta (increments) and obtained the top curve shown in Fig. 7. Except at the low end it is almost a straight line when plotted against decibels. The bottom curve is the sone scale of subjective loudness. The triangles represent the median results obtained by J. C. Stevens and E. Tulving (51) from 70 subjects who made direct magnitude estimations

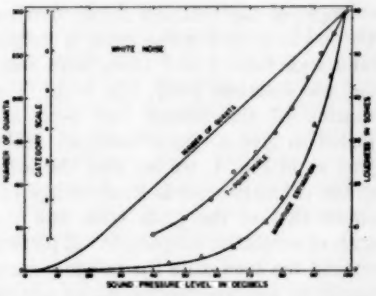


FIG. 7. J.n.d. (quantum) scale, category scale, and ratio scale (sones) for the loudness of white noise. *Circles:* with the spacing of the levels adjusted to give a "pure" category scale, 8 subjects each made two judgments of each level on a 7-point scale. The range was disclosed at the outset and the order of presentation was different for each subject. *Triangles:* median judgments by 70 subjects who made magnitude estimations of eight levels presented in irregular order in a classroom, with no designated standard. Estimates were transformed to a common modulus at the 75-db level, which was the first level presented.

of the loudness of different levels of white noise presented in irregular order. In a log-log plot these points fall close to a straight line like that shown in Fig. 6.

The curves in Fig. 7 show how vastly different are the two scales, the j.n.d. (or quantum) scale and the loudness scale. In this respect they confirm what we have already seen in Fig. 1. They also show that when we use the loudness scale to measure the subjective size of successive j.n.d.'s, their size grows rapidly as we go up the scale. Except near the low end, their growth approximates an exponential function. Actually, over the low end of the scale where the j.n.d. scale is curved there is a superficial similarity between the j.n.d. scale and the loudness scale, but over the major portion of the audible range the divergence between the two scales is large and obvious.

The middle curve in Fig. 7 shows an

example of the function obtained when the loudness of a white noise is judged on a scale from 1 to 7 (data from Stevens and Galanter [64], Fig. 9A). The spacing of the stimuli had been adjusted to give a "pure" category scale. Just as in Fig. 1, we see that the form of the category scale is intermediate between that of the j.n.d. scale and the scale of subjective magnitude. If plotted against the sone scale the category scale would be concave downward, as are all category scales on prothetic continua.

It has occasionally been asserted that some scales of sensory magnitude agree with the scale obtained by counting off j.n.d.'s (74). Hanes' (27) statement that this is true for brightness is a recent example, which has led, incidentally, to a certain amount of confusion. In a subsequent study Hanes (28) extended the range of his measurements far enough for us to see clearly that the scale of subjective brightness (in brils) departs from the j.n.d. scale in the expected way. Perhaps Hanes should have stated this explicitly.

I have summed the j.n.d.'s for brightness obtained by three different experimenters (2, 18, 35) and have compared the results with Hanes' bril scale. The pictures presented are essentially similar to that shown for loudness in Fig. 7, except that the curved tail at the lower end of the j.n.d. scale tends to be longer. The basic fact is that over the middle and upper ranges of intensity Weber's law holds fairly well for brightness, but the bril scale is a power function. Over the range of the fainter luminances, Weber's law breaks down to a degree sufficient to produce the apparent agreement Hanes spoke of. But it is quite wrong to jump to the general conclusion that j.n.d.'s add up to the bril scale.

We have already noted, in the discussion of Guilford's "nth-power law," that in order for the j.n.d.'s to add up to give the magnitude function, ΔS cannot in general be an increasing function of S. For most perceptual continua it must be a decreasing function of S. But this is "unheard of in psychophysical research." It would mean, for example, that it would be easier to tell when a centimeter was added to a meter than when a centimeter was added to a centimeter. Resolving power simply does not work that way.

Suppose now for a moment we risk Titchener's displeasure and try to think like a physicist. Physics also faces the problem of resolving power and in most physical measurements resolving power is relative—the accuracy of measurement is proportional to magnitude. This fact is expressed by saying that length can be measured to 1 part in 10^5, that resistance can be measured to 1 part in 10^6, or that time intervals can be measured to 1 part in 10^8. In this sense Weber's law has its analogue in physics. Now, it is conceivable that someone might propose that we give up our ordinary ways of measuring length, say, and devise instead a scale whose unit is the resolving power of our best procedures. We would then be doing what Fechner proposed. In terms of our present units, the new units of length would grow increasingly larger as we go up the scale, and they would do so for the same reason that the j.n.d.'s grow larger as we go up the psychological scale on a prothetic continuum. The laws of physics could be rewritten in terms of the new scales of units based on resolving power, but it would probably strike the physicist as less than sensible to try to do it—and not simply because it would be hard work. For reasons that he could readily justify, the physicist wants to start with magnitudes and not with resolving power. What I am suggesting is that there is sufficient parallel between physics and psychophysics to justify the same approach in both fields. In neither discipline should we try to

make resolving powers, variabilities, confusions, j.n.d.'s, discriminal dispersions or any such entities serve as units of magnitude on continua that behave like the prothetic continua of Class I.

SCALING THEORY

The same counsel recommends itself to those of us who devise scaling procedures for other than perceptual variables. As Gulliksen (26) complains, those who worry about psychophysical measurement often ignore the work on nonsensory scaling, and vice versa. Titchener (71) also once complained that psychologists have ignored the advances in subjective scaling achieved by the meteorologists. Insularity is probably inevitable in an age of specialization, but it has its dangers. One of the dangers is that the laws of human judgment developed under simple conditions, where we can better see what is going on, may be ignored when we work in more complicated realms where discernment is more difficult.

A large segment of psychological scaling theory has been built around models that make scale "units" out of some measure of "discriminal dispersion," but this is precisely the kind of unit that is demonstrably *not* invariant on prothetic continua. Variabilities and confusions provide a tempting starting point for psychological measurement, chiefly because dispersion among our judgments is something we always have with us, but the transforming of unreliability, inconstancy, or confusion into units of measure somehow needs more than assumptions to back it up. Our fundamental scales can, I think, stand on firmer ground than the mere scatter of our data.

It is recognized, of course, that measures of confusion often have a usefulness of their own, and that for certain purposes it is desirable to construct "equal discriminability" scales (16). The engineering of schemes for optimizing the coding of information on cathode-ray tubes presents an interesting example of how such measures may be used in the prediction and control of confusability (36). But our concern here is with a different problem, namely, whether measures of confusion can be used as units of magnitude.

The use of dispersion as a tool for magnitude scaling requires assumptions regarding its behavior. How does dispersion (in psychological units) vary over the psychological continuum in question? A variety of assumptions regarding the behavior of variability have been elaborated, but so far as I know, none of them come close to admitting the drastic growth in dispersion that we find to be true for Class I (prothetic) continua. Here, we recall, the dispersion (in subjective units) approaches an exponential function of the number of j.n.d.'s above threshold or, alternatively stated, it grows almost in direct proportion to the psychological magnitude. The assumption made in practice is usually that dispersion (in psychological units) is constant, or nearly so. This is like Fechner's assumption that j.n.d.'s are equal in subjective size. Direct experiments can sometimes verify this assumption on continua of Class II (metathetic), but on the "quantitative" continua of Class I they show that the assumption fails by a wide margin.

On continua that behave like Class I we would be closer to right if we began with the assumption that discriminal dispersion is not constant but is proportional to the psychological magnitude in question. When the psychological magnitude is a power function of the stimulus, this assumption is equivalent to saying that psychological values separated by equal units of dispersion on the stimulus scale stand in a constant ratio to each other. If

we were to start from this assumption, could we then proceed to construct a useful scale of the psychological magnitude from measurements made of discriminal dispersion? In principle we might erect a new type of scale, but we could not construct an interval or a ratio scale of the kinds we ordinarily employ. While working on the theory of scales (**54**) back in the late 1930's I had a hard time convincing myself that if we know concerning a series of values only that $a/b = b/c = c/d = \ldots$ we are powerless to proceed further toward what I have called a ratio scale of measurement. But the fact is that the equating of ratios gets us no further unless (a) we can also equate intervals, or (b) we can determine the numerical value of the ratio (as we do under the ratio methods reviewed above). An example of how we might use equated ratios *plus* equated intervals to construct a ratio scale is discussed elsewhere (**55**, p. 24).

But let us pursue for a moment the problem of what we might do when we have equated a set of ratios: $a/b = b/c = c/d. \ldots$ We can assume that we have an operation for ordering these values and that $a < b < c < d. \ldots$ The problem then is, how may we assign numerical values to this series? As is true when we have a series of equated intervals, $a - b = b - c = c - d = \ldots$, we can pick any two values arbitrarily (subject only to the ordinal requirement), and having done so, all the other values are determined. Thus for the ratio series, if $a = 2$ and $b = 6$ then $c = 18$, $d = 54$, etc. The next question is, what transformations are permitted on this scale? We can show that any value x may be replaced by x' provided $x' = kx^n$ where k and n are constants that may take on any positive values. In other words we can always transform the scale values by a power function. For example, if we square all the values, the ratios remain equal.

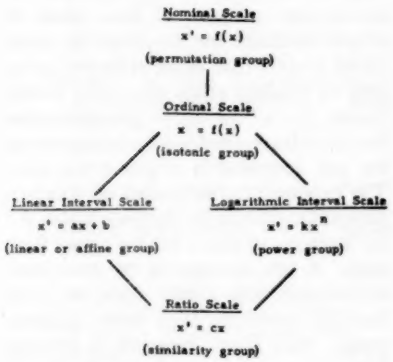
A scale of this kind may be mathematically interesting, but, like many mathematical models, it has thus far proved empirically useless. At least, so far as I am aware, no uses for it have yet been found. In a formal sense there are certain parallels between a scale of equal but indeterminate ratios and a scale of equal but indeterminate intervals. On both, for example, any two values can be selected arbitrarily, and the scale values of both can be transformed by an equation involving two arbitrary constants. But the scale of equal intervals has great utility, as witness the Fahrenheit and Celsius scales of temperature, and the scale of calendar time. And differences on interval scales can be measured on ratio scales, as when we measure intervals of duration in seconds. But what can we do with equated ratios when the value of the ratio is indeterminate?

An obvious suggestion is that we might convert to logarithms and express the equated ratios as $\log a - \log b = \log b - \log c = \log c - \log d$, etc. If we then restrict the values of a, b, c, . . . to positive numbers, we can set up an interval scale in logarithms. There is no a priori reason why we could not put this scale to work in a manner analogous to the workings of our linear interval scales. The linear and the logarithmic scales have in common the fact that the choice of the zero point is arbitrary. It might be interesting to explore the consequences of this procedure and perhaps try to answer the question why science has never made use of these logarithmic interval scales based on equated ratios.

Another question arises. If we were to develop such scales, what would we call them? When I originally proposed (**54**) the names *nominal, ordinal, interval,* and *ratio* for the four classes of scales commonly used, I worried some about the scale based on equated ratios,

but I dismissed it as a purely academic question, for there appeared to be no examples of its use. If uses for it were to develop, we would like, perhaps, to call this class of scales *ratio* scales, for they are based on an empirical operation for equating ratios. But since the term ratio scale has now gained wide currency as the name for the scales that admit only a similarity transformation (multiplication by a constant) it may be better to call the scale based on equated ratios a *logarithmic interval* scale.

With this class of scales included, the hierarchy of scales, together with the group of transformations permitted by each scale, would take the form shown in the diagram.



As this diagram suggests, a ratio scale is possible whenever empirical operations are available to create both types of interval scales. If we can determine both equal differences and equal ratios we can eliminate the additive constant b and the exponent n, and we are left with only a multiplying constant c. This defines a scale with a zero point—what we call a ratio scale, or, if we prefer, a "zero-point" scale.

Returning now to the assumptions made about discriminal dispersion we note that, in terms of the scales listed in the diagram, the assumption most often made is that the processing of variability will yield a linear rather than a logarithmic interval scale. It is not only in connection with the method of pair comparisons, which Thurstone called the "best of all the psychophysical methods" (70), that the wrong assumption of constant dispersion (in psychological units) often gets made. This assumption is either latent or explicit in many subtler procedures, such as the methods of "successive intervals," "graded dichotomies," "successive categories," etc. (see 23). The foregoing arguments suggest that whenever these methods are applied to prothetic continua, continua on which the j.n.d. (dispersion) tends to vary directly with magnitude, the resulting scales will turn out to be nonlinear. It seems obvious that none of the scales listed in Table 1 could be constructed by the method of pair comparisons, or by any of its several cousins. The laws of judgment derived from direct experiments on perceptual continua make it plain that, unless it can be demonstrated that the judgmental continuum behaves like those of Class II (metathetic), any scaling procedure that is geared to the assumption that equally often noticed differences are equal is defective.

Fortunately, this does not mean that no scaling is possible. We still have other devices, such as the direct methods for ratio scaling. Is there a genuine, nontrivial, substantive problem calling for a psychological scale for which these methods, or some variation on them, cannot produce an acceptable answer? Or is it merely that we like to assume, with Fechner, that indirection is the best path to the goal? Contrary to a common assumption, the use of the direct methods does not require knowledge of an underlying, measurable physical continuum (48). Only a nominal scale (55) is required at the stimulus level,

i.e., the stimuli must be identifiable by the experimenter. For example, when the observer judges the apparent ratio between two stimuli the experimenter needs to know which two stimuli were involved, but he does not need to know their values on any other scale. I know from experience that it is more comfortable to take it for granted that a direct method is impossible than it is to try to work one out. And it is still more onerous to test and refine the direct methods and to purge them of constraints and biases. The problem of the scaling of psychological continua is full of booby traps, but so is the laboratory measurement of a physical continuum like, for instance, electrical inductance or mechanical impedance.

Perhaps one of our professional idiosyncracies is that in psychology we are sometimes more enamored of models and methods than we are of problems. One occasionally gets the impression that there are more people with a method who are looking for a problem to use it on than there are searchers with a problem looking for a method. It is admittedly more entertaining to make a formal model, complete with assumptions, postulates, and derivations, than it is to grub through the empirical tangle of an experimental issue. But if something empirically useful is to issue from a model, something empirically known must be put into it. The gratuitous assumption that some measure of dispersion, sensitivity, or resolving power can be used as the unit of a scale of psychological magnitude does not meet this requirement.

SUMMARY

This paper has rambled far enough afield at this point and it is time to try to restate the issues in briefer compass. The main points are these.

Two general classes of perceptual continua can be distinguished by means of four functional criteria. On Class I or "quantitative" continua the j.n.d. increases in subjective size as psychological magnitude increases, category rating-scales are concave downward when plotted against psychological magnitude, comparative judgments exhibit a time-order error (a "category effect"), and equisection experiments exhibit hysteresis. On Class II or "qualitative" continua these four effects are apparently absent. Class I, called prothetic, includes those continua on which discrimination is mediated by an additive mechanism at the physiological level; Class II, called metathetic, includes those mediated by a substitutive mechanism.

On Class I (prothetic) continua the use of one or more of four kinds of direct methods for constructing ratio scales reveals that equal stimulus ratios tend to produce equal subjective ratios. Hence, to a first-order approximation the "psychophysical law" relating stimulus and response is a power function. The exponent, as measured on fourteen different continua, varies from about 0.3 for loudness to about 2.0 for visual flash rate. A few workers in the past have conjectured this power function, even starting sometimes from wrong assumptions. Only lately, however, is this law becoming securely anchored in experiment.

Fechner's logarithmic law is not found in experiment for the simple reason that resolving power (the j.n.d.) is not constant in psychological units, but is roughly proportional to psychological magnitude. For this reason, all procedures of Fechnerian extraction, like the method of pair comparisons and its related techniques, which seek to build scales out of "unitized" measures of dispersion, are not proper methods for scaling magnitudes that behave like Class I or prothetic continua. When-

ever psychological scales are called for, direct ratio scaling methods should probably be tried.

REFERENCES

1. BEEBE-CENTER, J. G., & WADDELL, D. A general psychological scale of taste. *J. Psychol.*, 1948, **26**, 517–524.
2. BLACKWELL, H. R. Contrast thresholds of the human eye. *J. opt. Soc. Amer.*, 1946, **36**, 624–643.
3. BORING, E. G. The stimulus-error. *Amer. J. Psychol.*, 1921, **32**, 449–471.
4. BRENTANO, F. *Psychologie vom empirischen Standpunkte*. Th. 1. Leipzig: Dunker & Hunblot, 1874.
5. COHEN, J., HANSEL, C. E. M., & SYLVESTER, J. D. Interdependence of temporal and auditory judgments. *Nature*, 1954, **174**, 642.
6. CORBIN, H. H., REESE, E. P., REESE, T. W., & VOLKMANN, J. Experiments on visual discrimination, 1952–55. Psychophysical Research Unit, Mt. Holyoke College, AFCRC TR No. 56–52, April 1956.
7. COWDRICK, M. The Weber-Fechner law and Sanford's weight experiment. *Amer. J. Psychol.*, 1917, **28**, 585–588.
8. EKMAN, G., BERGSTROM, B., & KÜNNAPAS, T. M. Comparison between two psychophysical scaling methods. Report from the Psychological Laboratory, Univer. of Stockholm, No. 37, August 1956.
9. EKMAN, G., & DAHLBÄCK, B. A subjective scale of velocity. Report from the Psychological Laboratory, Univer. of Stockholm, No. 31, February 1956.
10. FECHNER, G. T. *In Sachen der Psychophysik*. Leipzig, 1877.
11. FERNBERGER, S. W. On absolute and relative judgments in lifted weight experiments. *Amer. J. Psychol.*, 1931, **43**, 560–578.
12. Final Report. *Advancement of Science*, 1940, No. 2, 331–349.
13. FULLERTON, G. S., & CATTELL, J. McK. *On the perception of small differences*. Philadelphia: Univer. of Pennsylvania Press, 1892.
14. GARNER, W. R. Some statistical aspects of half-loudness judgments. *J. acoust. Soc. Amer.*, 1952, **24**, 153–157.
15. GARNER, W. R. A technique and a scale for loudness measurement. *J. acoust. Soc. Amer.*, 1954, **26**, 73–88.
16. GARNER, W. R., & HAKE, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
17. GILINSKY, ALBERTA S. Perceived size and distance in visual space. *Psychol. Rev.*, 1951, **58**, 460–482.
18. GRAHAM, C. H., & BARTLETT, N. R. The relation of size of stimulus and intensity in the human eye. III. The influence of area on foveal intensity discrimination. *J. exp. Psychol.*, 1940, **27**, 149–159.
19. GREGG, L. W. Fractionation of temporal intervals. *J. exp. Psychol.*, 1951, **42**, 307–312.
20. GROTENFELT, A. *Das Webersche Gesetz und die psychische Relativität*. Helsingfors: Frenckell, 1888.
21. GRUBER, H. E. The relation between perceived size and perceived distance. *Amer. J. Psychol.*, 1954, **67**, 411–426.
22. GUILFORD, J. P. A generalized psychophysical law. *Psychol. Rev.*, 1932, **39**, 73–85.
23. GUILFORD, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.
24. GUILFORD, J. P., & DINGMAN, H. F. A validation study of ratio-judgment methods. *Amer. J. Psychol.*, 1954, **67**, 395–410.
25. GUILFORD, J. P., & DINGMAN, H. F. A modification of the method of equal-appearing intervals. *Amer. J. Psychol.*, 1955, **68**, 450–454.
26. GULLIKSEN, H. Paired comparisons and the logic of measurement. *Psychol. Rev.*, 1946, **53**, 199–213.
27. HANES, R. M. A subjective brightness scale. *J. exp. Psychol.*, 1949, **39**, 438–452.
28. HANES, R. M. The construction of subjective brightness scales from fractionation data: a validation. *J. exp. Psychol.*, 1949, **39**, 719–728.
29. KARLIN, J. E., & STEVENS, S. S. Auditory tests for the ability to discriminate the pitch and loudness of noises. Report from the Psycho-Acoustic Laboratory, Harvard Univer., OSRD No. 5294 (PB 39576), 1 August 1945.
30. MACLEOD, S. A construction and attempted validation of sensory sweetness scales. *J. exp. Psychol.*, 1952, **44**, 316–323.
31. MERKEL, J. Die Abhängigkeit zwischen Reiz und Empfindung. *Phil. Stud.*, 1888, **4**, 541–594; 1889, **5**, 245–291; 1889, **5**, 499–557.
32. METFESSEL, M. F. A proposal for quan-

titative reporting of comparative judgments. *J. Psychol.*, 1947, **24**, 229–235.

33. MILLER, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *J. acoust. Soc. Amer.*, 1947, **19**, 609–619.
34. MOON, P. *The scientific bases of illuminating engineering.* New York: McGraw-Hill, 1936.
35. MUELLER, C. G. Frequency of seeing functions for intensity discrimination at various levels of adapting intensity. *J. gen. Physiol.*, 1951, **34**, 463–474.
36. MULLER, P. F., *et al.* Symbolic coding of information on cathode ray tubes and similar displays. *WADC Tech. Rep. 55–375.* Wright Air Development Center, Ohio, 1955.
37. MÜNSTERBERG, H. A psychometric investigation of the psychophysic law. *Psychol. Rev.*, 1894, **1**, 45–51.
38. NEWMAN, E. B. The validity of the just noticeable difference as a unit of psychological magnitude. *Trans. Kansas Acad. Sci.*, 1933, **36**, 172–175.
39. NEWMAN, E. B., STEVENS, S. S., & VOLKMANN, J. On the method of bisection and its relation to a loudness scale. *Amer. J. Psychol.*, 1937, **49**, 134–137.
40. PIÉRON, H. *The sensations, their functions, processes and mechanisms.* New Haven: Yale Univer. Press, 1952.
41. PLATEAU, M. H. Sur la mesure des sensations physique, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante. *Bull. de l'Acad. Roy. Belg.*, 1872, **33**, 376–388.
42. POLLACK, I. Auditory flutter. *Amer. J. Psychol.*, 1952, **65**, 544–554.
43. POSTMAN, L. The time-error in auditory perception. *Amer. J. Psychol.*, 1946, **59**, 193–219.
44. POSTMAN, L. Time-error as a function of the method of experimentation. *Amer. J. Psychol.*, 1947, **60**, 101–108.
45. POULTON, E. C., & STEVENS, S. S. On the halving and doubling of the loudness of white noise. *J. acoust. Soc. Amer.*, 1955, **27**, 329–331.
46. PRATT, C. C. Time-error in the method of single stimuli. *J. exp. Psychol.*, 1933, **16**, 798–814.
47. REESE, E. P. (Ed.) Psychophysical research summary report, 1946–1952. Psychophysical Research Unit, Mt. Holyoke College. USN Spec. Dev. Ctr. *Tech. Rep. No. SDC–131–1–5*, 1953.
48. REESE, T. W. The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples. *Psychol. Monogr.*, 1943, **55**, No. 3 (Whole No. 251).
49. RICHARDSON, L. F., & ROSS, J. S. Loudness and telephone current. *J. gen. Psychol.*, 1930, **3**, 288–306.
50. ROSS, S., & KATCHMAR, L. The construction of a magnitude function for short time intervals. *Amer. J. Psychol.*, 1951, **64**, 397–401.
51. STEVENS, J. C., & TULVING, E. Loudness estimation in a group situation. *Amer. J. Psychol.* (in press).
52. STEVENS, S. S. A scale for the measurement of a psychological magnitude loudness. *Psychol. Rev.*, 1936, **43**, 405–416.
53. STEVENS, S. S. On the problem of scales for the measurement of psychological magnitudes. *J. Unif. Sci.*, 1939, **9**, 94–99.
54. STEVENS, S. S. On the theory of scales of measurement. *Science*, 1946, **103**, 677–680.
55. STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
56. STEVENS, S. S. Pitch discrimination, mels, and Kock's contention. *J. acoust. Soc. Amer.*, 1954, **26**, 1075–1077.
57. STEVENS, S. S. On the averaging of data. *Science*, 1955, **121**, 113–116.
58. STEVENS, S. S. The measurement of loudness. *J. acoust. Soc. Amer.*, 1955, **27**, 815–829.
59. STEVENS, S. S. Decibels of light and sound. *Physics Today*, 1955, **8** (10), 12–17.
60. STEVENS, S. S. The direct estimation of sensory magnitudes: loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.
61. STEVENS, S. S. The calculation of the loudness of complex noise. *J. acoust. Soc. Amer.*, 1956, **28**, 807–832.
62. STEVENS, S. S. Concerning the form of the loudness function. *J. acoust. Soc. Amer.* (in press).
63. STEVENS, S. S., & DAVIS, H. *Hearing: its psychology and physiology.* New York: Wiley, 1938.
64. STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. Report from the Psycho-Acoustic Laboratory, Harvard Univer., PNR–186, August 15, 1956. (Submitted to *J. exp. Psychol.*)
65. STEVENS, S. S., MORGAN, C. T., & VOLK-

MANN, J. Theory of the neural quantum in the discrimination of loudness and pitch. *Amer. J. Psychol.*, 1941, **54**, 315–355.

66. STEVENS, S. S., & POULTON, E. C. The estimation of loudness by unpracticed observers. *J. exp. Psychol.*, 1956, **51**, 71–78.
67. STEVENS, S. S., & VOLKMANN, J. The relation of pitch to frequency: a revised scale. *Amer. J. Psychol.*, 1940, **53**, 329–353.
68. STEVENS, S. S., VOLKMANN, J., & NEWMAN, E. B. A scale for the measurement of the psychological magnitude pitch. *J. acoust. Soc. Amer.*, 1937, **8**, 185–190.
69. TAVES, E. H. Two mechanisms for the perception of visual numerousness. *Arch. Psychol., N. Y.*, 1941, No. 265.
70. THURSTONE, L. L. Fechner's law and the method of equal appearing intervals. *J. exp. Psychol.*, 1929, **12**, 214–224.
71. TITCHENER, E. B. The psychophysics of climate. *Amer. J. Psychol.*, 1909, **20**, 1–14.
72. TITCHENER, E. B. *Experimental psychology*. Vol. II, Part II (Instructor's Manual). New York: Macmillan, 1923.
73. WOODROW, H. Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Chap. 32.
74. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.
75. WRIGHT, W. D. *Researches on normal and defective colour vision*. St. Louis: Mosby, 1947.

(Received for early publication November 2, 1956)

Psychological Review
Vol. 64, No. 3, 1957

THEORY OF SELECTIVE LEARNING WITH PROBABLE REINFORCEMENTS

FRANK RESTLE [1]

Center for Advanced Study in the Behavioral Sciences [2]

This paper presents a quantitative description of some simple selective-learning data. The descriptive system is shown to subsume the writer's theory of discrimination learning (**11**), and to relate closely to Estes' statistical theory (**5**).

Recent mathematical formulations of learning (**2, 3, 4, 5, 6, 11**) describe the stimulus situation as a set of elements, each of which is conditioned to (tends to evoke) exactly one response at a given time. During learning, if a certain response A_1 is reinforced, a cue may switch and become newly conditioned to A_1. The probability of such a change is the rate-of-learning parameter. Estes (**4, 5, 7**) and Bush and Mosteller (**2, 3**) have shown that such a model has many consequences for learning with random reinforcement schedules, correctly predicting asymptotes in guessing experiments (**3, 5, 7**). However, certain discrepancies have appeared when theoretical expectations are compared with performance on early trials. Estes[3] has listed some problems which, if solved, would contribute to a workable descriptive theory of elementary learning. One of these is the tendency of the parameter representing learning rate to vary systematically with reinforcement probabilities. In general, learning is more rapid when the probabilities of reinforcement are more extreme. A second problem is that subjects can master discrimination problems to perfection even in the presence of "overlap" stimuli which are not differentially reinforced. The writer's discrimination theory (**11**) has some success in resolving the stimulus-overlap problem by assuming that irrelevant cues are adapted. This paper extends this discrimination theory to cases of random reinforcement, using Estes' equations for probability learning, but, in addition, offering a descriptive account of how learning rate may depend on reinforcement probability.

The major points to be made in this paper are that (1) the stimulus situation in two-choice learning experiments is represented by a set of discriminable aspects called "cues"; (2) a cue may be "conditioned" to either response; (3) a cue may be "adapted" and rendered nonfunctional during learning; (4) the probability of a response is the proportion of the unadapted cues conditioned to it; (5) the degree to which a cue is differentially reinforced determines its "validity"; and that (6) the rates of conditioning and adaptation depend on the validities of the cues available.

When these assumptions are stated quantitatively, the theory is shown to have three characteristics: first, under certain limiting experimental conditions it leads to a process of probability learning similar to that in Estes' theory, and in particular with the same

[1] Appreciation is due Dr. W. K. Estes for his advice and criticism, and for lending prepublication drafts of papers. The writer has benefited from extended seminars and conversations with L. B. Wyckoff, P. Suppes, and D. H. Howes.

[2] Now at Michigan State University.

[3] See Estes, W. K., The statistical approach to learning theory (draft), prepared for the report of APA "Project A," *Analysis of psychological science.*

asymptotes; second, when applied to discrimination learning it has the same consequences as the writer's earlier discrimination-learning theory (**11**), with the same way of handling stimulus overlap; and third, it describes the connection betweeen reinforcement probabilities and learning rates in the simplest probability-learning experiments.

THEORY

The Set of Cues

We think of the stimulus situation as composed of a set of discriminable aspects, called "cues," k, k', k''. The set of cues is called K and the number of cues is N. The proportion of cues belonging to any manipulated aspect of the situation is measured from behavior, and the number of cues, N, is for most purposes arbitrary.

Conditioning of Cues

Following Estes (**4**), we assume that a cue is conditioned to one or the other response alternative at any time, on an all-or-none basis. The probability that cue k is conditioned to response A_1 at trial n is called $F(k,n)$. If a cue is conditioned to A_2 and then A_1 is reinforced, it may switch over and become conditioned to A_1. The probability of such a switch is a constant called θ, the "rate-of-learning" parameter. On this assumption, we get the following equations of change of $F(k,n)$:

i) If A_1 is reinforced on trial n,

$$F(k,n+1)=F(k,n)(1-\theta)+\theta. \quad (1)$$

ii) If A_2 is reinforced on trial n,

$$F(k,n+1) = F(k,n)(1-\theta). \quad (2)$$

If A_1 is reinforced on some fixed proportion π of the trials and A_2 is reinforced on all other trials, the mean equation is

$$F(k,n+1)=F(k,n)(1-\theta)+\pi\theta \quad (3)$$

which has the solution

$$F(k,n_0+n) = \pi - [\pi - F(k,n_0)](1-\theta)^{n-1}. \quad (4)$$

Equation 4 is the conditioning equation which will be used in the applications discussed in this paper.

Adaptation of Cues

During learning, a cue may become "adapted" (**11**) and lose its effect on response. Different cues have different probabilities of becoming adapted. If cue k is not adapted by the beginning of trial n, the probability that it will be adapted by trial $n+1$ is $\phi(k)$. This consideration gives us an equation for $a(k,n)$, the probability that cue k is adapted at the beginning of trial n, as follows:

$$a(k,n+1) = a(k,n) \times [1-\phi(k)] + \phi(k), \quad (5)$$

which has the solution

$$a(k,n_0+n) = 1 - [1-a(k,n_0)] \times [1-\phi(k)]^{n-1}. \quad (6)$$

Equation 6 is the adaptation equation which will be used in applications.

Probability of a Response

The probability of response A_1 is the proportion of unadapted cues conditioned to it. The probability that cue k is unadapted is $1-a(k,n)$, and the probability that it is conditioned to A_1 is $F(k,n)$. Thus,

$$p(n) = \frac{\Sigma F(k,n)[1-a(k,n)]}{\Sigma[1-a(k,n)]}, \quad (k \text{ a member of } K). \quad (7)$$

Validity of Cues

The experimenter determines the validity of a cue by differential reinforcement. Suppose that one response to cue k is reinforced proportion π of the time and the other is reinforced $1-\pi$ of the time. The cue is a good predictor of reward if π is near one or zero, and it is useless if $\pi = .5$. The term "validity" underscores an analogy between the problem facing an S in selective learning and that facing a diagnostician. As the diagnostician tries to predict some criterion, using various test items with varying validities, so the S tries to select the correct response, using various cues which may be related in varying degrees to reward.

Validity of a cue, V, may be measured in several ways, but to get the simplest empirical laws it seems best to define it as

$$V = 4\pi^2 - 4\pi + 1. \qquad (8)$$

The use of a quadratic function is suggested by the possibility that validity depends on the proportion of pairs of trials on which consistent information is received, $\pi^2 + (1-\pi)^2$, and on which contradictory information is received, $2\pi(1-\pi)$. The coefficients of the quadratic equation are fixed by making the validity of a consistently reinforced cue ($\pi = 1$ or $\pi = 0$) equal to one, and the validity of an irrelevant cue ($\pi = .5$) equal to zero.

Rates of Conditioning and Adaptation as Functions of Cue-Validity

In the writer's discrimination learning theory (11) it was assumed that the rates of conditioning of relevant cues, θ, and adapting irrelevant cues, $\phi(k')$, were equal to one another and were also equal to the *proportion of relevant cues*. This basic idea is now generalized to experiments with various reinforcement probabilities and to cues with validities between zero and one.

It is assumed that the rate-of-learning parameter controlling conditioning of cues, θ, is equal to the mean validity of cues. That is,

$$\theta = \Sigma\, V_k/N, \quad (k \text{ a member of } K). \qquad (9)$$

If all cues have validities 1 or 0, as is the case in the discrimination theory, then the mean validity is equal to the proportion of relevant cues.

It is assumed that the probability of adapting cue k' depends on the number of cues more valid than k', and the degree to which they are more valid. Mathematically

$$\phi(k') = \Sigma\, (V_k - V_{k'})/N, \quad (k, k' \text{ in } K, \text{ and } V_k \geq V_{k'}). \qquad (10)$$

If $V_{k'} = 0$, then $\phi(k') = \theta$ and the probability of adapting an invalid cue is the same as the probability of conditioning. If there is no cue with higher validity than k', then $\phi(k') = 0$ and k' is not adapted.

Relation to Statistical Learning Theory

It was stated that the present theory is, under certain circumstances, very like Estes' statistical learning theory. If all relevant cues have the same probability, $F(k,n)$, of being conditioned to the A_1 response, and if all irrelevant cues are adapted, then from Equation 9

$$p(n) = F(k,n). \qquad (11)$$

If we substitute $p(n)$ for $F(k,n)$ in the difference Equations 2 and 3, we have the simplest cases of statistical learning theory. From Equation 7 we see that as n increases, $a(k',n)$ asymptotically approaches 1, so that after sufficient training all irrelevant cues will be

adapted. Thus as training progresses the present theory gets closer and closer to the Estes model, and the asymptotes of the Estes model are in at least some cases the asymptotes of the present theory. In analysis of the present cue-learning theory it is often convenient, if only asymptotic probabilities are of interest, to study Estes' simpler model.[4]

This correspondence between the present theory and statistical learning theory is neither an accident nor a convergence of independent theoretical approaches. The elegance and empirical success of the Estes and Bush-Mosteller formulations led the present writer to adopt their fundamental notions and formulas, and the correspondence to earlier models was built into this theory.

Relation to Discrimination Learning Theory

In an earlier paper (11) the writer offered a theory of two-choice discrimination with consistent reinforcement of relevant cues. A discrimination learning problem was described as having r relevant cues (with validity $V_k = 1$) and i irrelevant cues with validity $V_{k'} = 0$. At Trial 1 neither relevant nor irrelevant cues are differentially conditioned—a situation we may represent in the present notation as

$$F(k,1) = F(k',1) = .5. \quad (12)$$

Similarly, no cues are adapted at Trial 1, so that

$$a(k,1) = a(k',1) = 0. \quad (13)$$

From Equation 9,

$$\theta = r/(r + i) \quad (14)$$

and from Equation 10,

$$\phi(k) = 0 \quad (15)$$

$$\phi(k') = \theta. \quad (16)$$

These bounding conditions inserted in Equations 4, 6, and 7 yield the equations of discrimination learning theory.

Validity and the Rate of Learning in Guessing Experiments

Several experiments on guessing as a function of the probability of reinforcement have found that the rate-of-learning parameters depend on π. Our notion of cue-validity, and the assumption that learning rates depend on validity (Equations 9 and 10), are intended to give a quantitative description of such findings.

In each such experiment we assume that there are r relevant cues with reinforcement probability π, and i irrelevant cues with reinforcement probability .5. Let k be a typical relevant cue, and k' a typical irrelevant cue. Substituting Equations 4 and 6 into Equation 7 and noting that $\phi(k') = \theta$, we have

$$p(n_0+n) = \frac{r\{\pi-[\pi-F(k,n_0)](1-\theta)^{n-1}\}+iF(k',n_0+n)[1-a(k',n_0)](1-\theta)^{n-1}}{r+i[1-a(k',n_0)](1-\theta)^{n-1}}. \quad (17)$$

If we let $\delta = r/(r + i)$ be the proportion of relevant cues, it follows from Equation 11 that

$$\theta = \delta V \quad (18)$$

where V is defined by Equation 8.

[4] An important advantage of Estes' and Bush and Mosteller's theories is their mathematical development, which gives detailed statistics of data in a wide variety of random reinforcement situations. The equations for $p(n)$ of the present theory are considerably more difficult to work with in most cases.

If we divide all the terms in the right member of equation 17 by $r + i$ and use the definition of δ to simplify, and at the same time substitute δV for θ, it follows that

$$p(n_0+n) = \frac{\delta\{\pi - [\pi - F(k,n_0)](1-\delta V)^{n-1}\} + (1-\delta)F(k',n_0+n)[1-a(k',n_0)](1-\delta V)^{n-1}}{\delta + (1-\delta)[1-a(k',n_0)](1-\delta V)^{n-1}}. \quad (19)$$

Inspection of Equation 19 shows that if $(1 - \delta V)^{n-1}$ approaches zero the limit of $p(n_0 + n)$ is π. In such a situation the subject approaches the "probability matching" asymptote, eventually making response A_1 as frequently as it is reinforced. The larger V is, the more rapidly will this asymptote be approached, for the larger V the smaller $(1 - \delta V)^{n-1}$. Thus Equation 19 relates the speed of learning to validity, and through validity to probability of reinforcement, in an appropriate fashion.

The aim of this paper is to make accurate quantitative statements about the relationship of learning rate to V. To do this we must establish the starting conditions by determining for some trial, n_0, the values of $F(k,n_0)$, $F(k',n_0)$, and $a(k',n_0)$.

Starting Conditions in Guessing Experiments

In the experiments that we wish to consider, subjects commonly show a noticeable bias, so that $p(1) \neq .5$. It is possible to estimate $p(1)$ from the proportion of A_1 responses made on early trials. We do not know which cues are responsible for any bias that appears, so for convenience we assume that the initial bias is distributed equally over all cues. That is, we assume as a starting condition that

$$F(k,1) = F(k',1) = p(1). \quad (20)$$

Besides initial biases, one further consideration arises in regard to starting conditions. While the experimenter may *intend* to reinforce alternative A_1 on some fixed proportion π of the trials, he can actually accomplish this only on the average. On the first trial he must reinforce one response or the other, and the actual relative frequency of reinforcements of A_1 becomes either 1 or 0. It may happen that the subject gets only reinforcements of A_1 and no reinforcements of A_2 for the first four or six or even ten trials, though some A_2 reinforcements are to be given. During this initial block of trials the problem given the subject is just the same as if the experimenter intended to give reinforcement consistently with $\pi = 1$. Thus, for some initial block of n_0 trials, the subject works on a problem with actual $\pi = 1$ and $V = 1$, and then shifts over to a problem with intermediate reinforcement probabilities.

An exact learning curve $p(n)$ can be written using the actual sequences of reinforcements. Computations are easier if we assume that, for the first $n_0 = \pi/(1 - \pi)$ trials, only A_1 reinforcements are given ($\pi > .5$), on Trial $n_0 + 1$ an A_2 reinforcement is given, and on all subsequent trials π is a satisfactory approximation to the actual relative frequency of A_1 reinforcements.

Probability of an A_1 *Response on Trial* n *in a Guessing Problem*

During the first n_0 trials the subject has a problem just like simple discrimination learning with starting probability $p(1)$. Thus for any n' less than or equal to $n_0 + 1$,

$$p(n') = \frac{\delta\{1 - [1 - p(1)](1 - \delta)^{n'-1}\} + (1 - \delta)p(1)(1 - \delta)^{n'-1}}{\delta + (1 - \delta)^{n'}}, \quad (21)$$

whereas for any later trials $n_0 + n$ we employ Equation 19 with

$$F(k,n_0 + 2) = \{1 - [1 - p(1)] \times (1 - \delta)^{M_0}\}(1 - \delta)$$

$$F(k',n_0 + 2) = .5 - (.5 - p(1)) \times (1 - \delta)^{n_0+1}$$

$$a(k',n_0 + 2) = 1 - (1 - \delta)^{n_0+1}.$$

In this fashion we construct a theoretical learning curve for learning in a guessing experiment.

In order to estimate parameters and test the theory conveniently, we wish to avoid the laborious process of fitting Equations 19 and 21. To do this we use the theoretical "total score" (as in **11**, Equation 8), obtained by cumulating the learning function $p(n)$. The method, accurate to a good approximation, is to replace the discrete function $p(n)$ by a function of a continuous variable t, and integrate the resulting $p(t)$. The integration is elementary[5] and gives an explicit function which, while bulky, is usable. We shall call this integral (which is not given here) T_m representing the cumulative $p(n)$ from 1 to m. T_m is the total number of A_1 responses made on the first m trials of training.

Empirical Estimates of V *as a Function of* π

If two or more groups of subjects from the same population are trained with the same apparatus and instructions but with different reinforcement probabilities, we assume they have the same cues available, and thus the same proportion of relevant cues δ, though the validity of the relevant cues will differ. In such an experiment one group of subjects can be used to estimate δ, and this same δ can be used for the other group, permitting a direct numerical estimate of V as a function of π. Grant, Hake, and Hornseth (**9**) ran five groups of twenty subjects each in a guessing experiment with reinforcement probabilities of 1.00, 0.75, 0.50, 0.25, and 0.00. Symmetry in the data permits us to lump the 1.00 and the 0.00 groups, calling them $\pi = 1$, and the .75 and .25 groups, calling them $\pi = .75$. The .50 group showed virtually no learning, as expected, and is disregarded.

The data from this experiment are used to check the assumed validity function given in Equation 8. The relevant cues for the $\pi = 1$ group were assigned validity $V = 1$. From this group we estimated the proportion of relevant cues, δ. Using this value δ, we then found the value of V which would cause the .75 group to make 34.5 A_1 reponses in 60 trials, as observed. The results of the computations are shown in Table 1. The obtained value of V at .281 is relatively close to the value predicted by Equation 8, which is $4(.75^2) - 4(.75) + 1 = .250$. There seems to be no way to test the statistical reliability of the discrepancy.

A second measure of V is accomplished by using the data from an experiment by Goodnow and Postman (**8**). They used reinforcement probabilities of 1.0, .9, .8, .7, .6, and .5 in a "problem-solving" situation. A peculiarity of their technique was a pretraining session in which the subjects were instructed and trained to attend to the relevant cues. We shall assume that this pretraining had the same effect as adapting all irrelevant

[5] See Burington's Tables (1, p. 82), formulas 309 and 310.

TABLE 1

SUMMARY OF CALCULATIONS ON GRANT–HAKE–HORNSETH EXPERIMENT

π	$p(1)$	$n_0=\pi/(1-\pi)$	δ	T_∞	Estimated V	Predicted V
1.00	.46	—	.137	52.7	—(Assumed to be 1.0)—	
.75	.46	3	.137	34.5	.281	.250

cues before the actual training trials began. The assumption is stated as

$$a(k',1) = 1$$

for any irrelevant cue k'. With this assumption we see that, according to Equation 7, for all trials n,

$$p(n) = F(k,n).$$

With this special assumption it is possible to sum $p(n)$ directly, dispensing with the approximation.

In analysis of the Goodnow-Postman experiment we again use the $\pi = 1$ group to estimate δ, the proportion of relevant cues. For all other groups all parameters except $p(1)$ were known, and estimates of $p(1)$ were made easily, since all groups showed about the same bias. Under these circumstances it is possible to make empirical estimates of V as a function of reinforcement probability. This was done using the .9, .8, .7, and .6 groups. The detailed results of the computations are shown in Table 2. The correspondence between estimated and predicted values of V is quite good, especially for those groups which learned slowly and had small values of V. Since there were only eight subjects per group the estimates of V are probably not highly accurate. However, when the estimates of V from the Grant-Hake-Hornseth and Goodnow-Postman experiments are plotted against the prediction from Equation 8, as in Figure 1, the results are encouraging.

If we accept Equation 8 as representing the relationship between validity (and thus learning rate) and π, the theory can be used to explain the detailed results of guessing experiments reported by Estes and Straughan (**7**) and Jarvik (**10**). Estes and Straughan trained 16 subjects in each of two groups with reinforcement probabili-

TABLE 2

SUMMARY OF CALCULATIONS ON GOODNOW–POSTMAN EXPERIMENT

π	$p(1)$	$n_0=\pi/(1-\pi)$	δ	T_∞	Estimated V	Predicted V
1.0	.43	—	.078	73.7	—(Assumed to be 1.00)—	
.9	.43	9	.078	66.7	.860	.640
.8	.43	4	.078	54.5	.260	.360
.7	.43	2	.078	48.0	.161	.160
.6	.43	2	.078	42.0	.041	.040

TABLE 3

SUMMARY OF CALCULATIONS ON THE ESTES–STRAUGHAN AND JARVIK EXPERIMENTS

Estes-Straughan Experiment

π	$p(1)$	n_0	δ	Theoretical V	Total A_1 Responses (Obtained)	Total A_1 Responses (Theoretical)
.85	.50	6	.167	.490	95.8	—
(.70)	.42	2	.167	.160	67.3	67.8

Jarvik Experiment

π	$p(1)$	n_0	δ	Theoretical V	Total A_1 Responses (Obtained)	Total A_1 Responses (Theoretical)
.744	.46	3	.342	.238	60.0	—
.663	.46	2	.342	.106	52.0	49.8
.593	.46	2	.342	.035	46.5	46.8

ties of .85 and .30. Our computation is to estimate δ, the proportion of relevant cues, from the data of the .85 group, and then assuming that δ is the same for the .30 group and V depends on π according to Equation 8, compute the theoretical performance of the .30 group. Jarvik had three groups with reinforcement probabilities of .744, .663, and .593. The data of the .744 group are used to estimate δ, and then performance curves are predicted for the other two groups according to the theory. The detailed results of these computations are shown in Table 3. Inspection of the results shows that the theory is able to predict the total A_1 reponses in m trials with acceptable accuracy in these two experiments.

Figure 2 summarizes the empirical accuracy of the proposed theory. The group with the highest reinforcement probability in each experiment is used to predict the results of the other groups, and theoretical expectations are compared with observed data. In each of the predictions the starting probability $p(1)$ was estimated from the data of the group predicted. Other than this, the predictions are computed without reference to the performance of the group predicted.

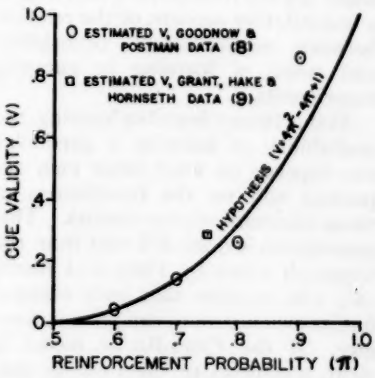


FIG. 1. Cue validity as a function of reinforcement probability—theoretical hypothesis (Eq. 8) and empirical estimates.

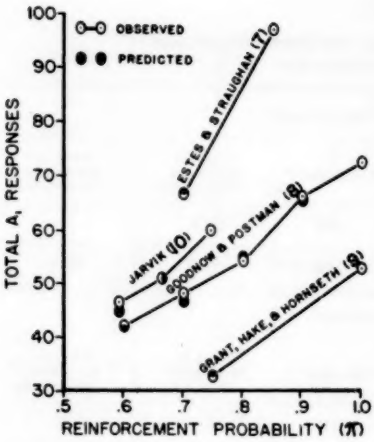


FIG. 2. Theoretical predictions of total-response scores in four experiments with empirical observed means.

DISCUSSION

The theoretical formulation presented in this paper has been shown to give rise to the same asymptotes as does Estes' theory in simple probability-learning situations, to have the same consequences as the writer's earlier discrimination learning theory when applied to such data, and to give a quantitative account of the relation between reinforcement probability and speed of learning in guessing experiments.

As this theory describes learning, the probability of learning a particular cue depends on what other cues are present and on the correlations of those cues with reinforcements. This assumption is quite different from the approach taken by Estes and Burke (6), who suppose that each element has its own invariant sampling parameter. If the Estes-Burke model is right, a complex problem can be predicted theoretically as an assembly of stimulus elements, each with its own fixed parameter. Using the present theory, we consider a complex problem as an assembly of elements which may appear in simpler problems, but the parameters of these elements depend on the context of the problem as a whole. While analysis of a problem as a whole seems at first too difficult, the results reported in this paper indicate that appropriate simplifying assumptions may make definite computations feasible.

The empirical results show that the present theory correctly describes the relation of learning rate to reinforcement probability. These results do not constitute evidence that the theory is correct, since the theory was formulated in the light of those results. In particular, the function relating validity to reinforcement probability, Equation 8, was selected because it seemed to fit the data. It is hoped that this formulation will stimulate the collection of more precise parametric data regarding the rate of learning under conditions of probable reinforcement.

REFERENCES

1. BURINGTON, R. S. *Handbook of mathematical tables and formulas.* (3rd ed.) Sandusky, Ohio: Handbook Publishers, 1955.
2. BUSH, R. R., & MOSTELLER, F. A model for stimulus generalization and conditioning. *Psychol. Rev.*, 1951, **58**, 413–423.
3. BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.
4. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
5. ESTES, W. K. Theory of learning with constant, variable, or contingent probabilities of reinforcement. In press, *Psychometrika.*
6. ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
7. ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

8. GOODNOW, J. J., & POSTMAN, L. Probability learning in a problem-solving situation. *J. exp. Psychol.*, 1955, **49**, 16–22.

9. GRANT, D A., HAKE, H. W., & HORNSETH, J. F. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

10. JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

11. RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

(Received August 23, 1956)

Psychological Review
Vol. 64, No. 3, 1957

VALUE AND THE PERCEPTUAL JUDGMENT OF MAGNITUDE [1]

H. TAJFEL

University of Durham, England [2]

Problems connected with the influences exercised by motivational factors on perception have been subject to much debate in recent years. It seems that the dust of controversy is beginning to settle. A reaction is setting in (e.g., **17**) against premature attempts to construct theories which would cope with the enormous variety of facts in this field, even before many of the facts have been properly established. These general theories were based in part on an a priori assumption that all the phenomena concerned can be reduced to, and formulated in terms of, a few very wide and nonspecific principles. No one can tell at present whether such an assumption is justified; but its hasty acceptance seems to have led to a confusion of experimental issues which has not been avoided either by those who stressed the importance of motivational variables in perception or by those who persistently denied it. The primary concern of this paper is with only one of these issues: the phenomena of perceptual overestimation, presumably due to value or some related variables. One of the reasons for this concern is that there exists a body of positive experimental findings in this field which cannot be ignored, and for which a consistent explanation has not yet been offered.

In the last fifteen years or so about twenty experiments on various aspects of overestimation have been conducted. Of those, only two have yielded unambiguously negative results (**3**, **21**). Partly negative results have been reported by Carter and Schooler (**8**), and by Klein, Schlesinger, and Meister (**19**). All other workers were able to conclude that, in the situations which they were using, "motivational" or "value" variables had an effect on their subjects' perceptual judgments of magnitude. Shifts in judgments of size (e.g., **1**, **2**, **5**, **6**, **7**, **10**, **20**, **26**), weight (e.g., **9**), number (e.g., **25**), and brightness (**13**) have been reported.

All this evidence cannot be dismissed as an accumulation of experimental artifacts. Furthermore, perceptual accentuation need not be considered as a "maladaptive" phenomenon. The fact that it may represent a departure from the "objective reality" leads to a general criticism which consists in pointing out that in order to survive we must perceive the world as it is, that we usually do, and that therefore the fleeting phenomena of overestimation are more typical of the specific laboratory situations in which they have been demonstrated than of perception under normal conditions. There is, however, a possibility that the shifts in the judgments of magnitude which concern us here do not interfere with an adequate handling of the environment. They may even be of help.

A Classification

The experiments on overestimation fall naturally into two classes. In one group, changes in the magnitude of the stimuli under investigation are relevant to the changes in value. The experiments on coins provide an example here: in general, the larger the coin, the greater

[1] I wish to express my gratitude to Professor J. S. Bruner for encouragement and advice; and to Dr. W. Sluckin and Mr. D. Graham from the University of Durham for many useful suggestions.

[2] Now at the University of Oxford, Delegacy for Social Training.

its value. The experiment of Dukes and Bevan (9), in which judgments of weight of jars filled with sand were compared with judgments of weight of jars filled with candy, would also be "relevant" in this sense: heavier jars would presumably contain more candy, and thus have greater "value." On the other hand, several experiments have been reported in which changes in value have in no apparent way been related to changes in the physical dimension which the subjects were requested to judge. Thus, in the experiment by Lambert, Solomon, and Watson (20), the color of the disc was the determinant of its value, as red discs only were associated with reward; but judgments of size were requested of the subjects. In the experiments by Bruner and Postman (6) and by Klein, Schlesinger, and Meister (19), discs containing a swastika were among the stimuli used. Judgments of their size were compared with judgments of size of discs containing neutral symbols. Here again, the size of the swastika has no easily conceivable relationship to its degree of relevance. Further examples from both classes of experiments could be given.

"RELEVANT" DIMENSIONS AND THE ACCENTUATION OF DIFFERENCES

Let us first deal with the "relevant" group of experiments. The stimuli whose magnitudes were judged in these investigations form, by definition, a series varying concurrently in at least two dimensions: the physical dimension (size or weight), and the "dimension" of value. The concern of most experimenters has been to show (or to deny) that the stimuli of the series in which a variation in magnitude was paired in the environment with a variation in value were judged larger than stimuli of objectively equivalent magnitudes belonging to a different, "neutral," series. Thus, judgments of size of coins were compared with judgments of size of cardboard or metal discs, and judgment of weight of jars filled with candy with similar judgments of jars filled with sand. Little attention has been paid to the perceived differences *between* the magnitudes of the stimuli in the "valued" series, as compared with the corresponding relationships in the "neutral" series.

It seems that some, at least, of the apparent contradictions between the various experimental results can be resolved if this intraserial aspect of the situation is considered. On the basis of an argument to be developed later, a prediction can be made that in a "relevant" series, where value changes concurrently with the dimension subjected to investigation, the differences between the stimuli of the series will be perceived as larger than the objectively equivalent differences between the stimuli of a neutral series, where no such association exists between value and magnitude. This comparison of the perceived differences within the two series would acquire its fullest meaning when intraindividual data are considered. Large individual variations are to be expected in the fineness of discrimination for both relevant and neutral series, but the tendency for a larger accentuation of differences in the former is likely to appear in most subjects, independently of the absolute values of the differences.

All the results to date in the "relevant" group of experiments support this contention. The quantitative difference, as expressed by subjects' judgments, between the two extremes of a series has been adopted as a rough measure of the accentuation of differences. This limitation was found necessary for two reasons. In the first place, coins, for example, differ not only in size and value. Secondly, it is possible that other determinants of quantitative judgment, such as the interaction of various gen-

TABLE 1

(CONVERTED FROM CARTER AND SCHOOLER, P. 202)*

Stimuli	Percentage Increase from Actual to Perceived Extension
Cardboard discs	16.81
Aluminum discs	16.20
Coins judged when present	33.91
Coins judged from memory	38.69

* Judgments of the "dime" and of the corresponding discs in other series have not been taken into account in this table, as the "dime" represents an exception in the concomitant variation of size and value in coins.

TABLE 2

(CONVERTED FROM BRUNER AND RODRIGUES, TABLE 6, P. 21)

Stimuli	Percentage Increase from Actual to Perceived Extension	
	Value Set	Accuracy Set
Coins (on table)	68.2	73.7
Metal discs	34.9	55.7
Paper discs	20.9	39.8

eralization gradients (cf., e.g., **18**), might counteract in some segments of the series the expected accentuation of differences.

A glance at Fig. 1 in the early study by Bruner and Goodman (**5**) shows that the perceived difference between the smallest and the largest coins (extension of the scale of judgments) is much larger than the corresponding extensions for discs. A conversion of the data in the experiment by Carter and Schooler (**8**, Table I, p. 202) to similar extensions, for all subjects combined, shows that:

a. In all series of stimuli (coins, aluminum and cardboard discs) the perceived extensions are larger than the actual ones;

b. The relative differences between the actual and the perceived extensions are about twice as large for the series of coins as for both series of discs (see Table 1).

Carter and Schooler imply this when they " . . . suggest that there is a constant error involved in making these size judgments such that small coins are underestimated and large coins are overestimated in size" (**8**, p. 205); but they do not seem to draw further conclusions.

Bruner and Rodrigues (**7**) have introduced in their well-controlled experiment the notion of "relative increase in overestimation" in which some of the suggestions made here are already implicit. Half of their subjects were assigned to a "value set" in which the instructions to the subjects emphasized the purchasing power of money. The other half were assigned to an "accuracy set" in which the experimenters' concern with accurate judgments of size was stressed. A conversion of their data similar to the above conversion of Carter and Schooler's data gives similar results (see Table 2).

This table is suggestive in several ways:

a. The same general tendency as in Carter and Schooler's data emerges: the relative extensions for the value series are considerably larger than for both neutral series.

b. All extensions in the "accuracy set" are larger than the corresponding extensions in the "value set." This is not inconsistent with the general proposition put forward here that the relevance of a series of stimuli is related to an emphasis of the differences between them. It is reasonable to assume that an appeal for accuracy of judgment would lead to similar results in its stress on clear discrimination between the stimuli of the series.

c. The differences between the extensions of the value and the neutral series

are less marked in the "accuracy" than in the "value" set. This is also consistent with the position adopted in this paper. An appeal for accuracy would have little effect in a series associated with value, where the accentuation of differences between the stimuli exists already as a long-standing product of past experience (cf. McCurdy's discussion of the role of schemata in the perception of coins, **22**). This same appeal would have comparatively more effect in the case of a neutral series.

One further example: Dukes and Bevan (**9**) found that for their "positive" series (jars filled with candy) the variability of responses was less marked than for the "neutral" series (jars filled with sand). This is related to smaller j.n.d.'s in the positive than in the neutral series. In other words, the differences between the stimuli of the positive series were more clearly and consistently perceived than the objectively equivalent differences in the neutral series, and the scale of judgments was more extended for the positive than for the neutral series.

This summary of evidence suggests that the results of the "relevant" experiments on overestimation cannot be solely the product of some simple and rather mysterious process of overestimation. As has already been pointed out, the interest of the experimenters was confined mainly to the comparisons of the perceptual judgments of stimuli in a value series with the judgments of physically equivalent stimuli in a neutral series. The subjects were invited to make comparisons between the valued stimuli and either a neutral standard or elements of a parallel neutral series (with the exception of the experiment by Dukes and Bevan, in which comparisons *between* the stimuli of the value series were also included). However, the implications of the fact that the stimuli associated with value do form a series cannot be ignored; during the experiments the subjects were repeatedly exposed to the various elements of this series, and the "belongingness," in the case of coins, would be further enhanced by familiarity.

In view of the evidence concerning the effects that all elements of a series, past and present, have on the quantitative judgments of its individual members (e.g., **14**, **18**), it may reasonably be assumed that the judgments of magnitude given by the subjects in the "relevant" overestimation experiments were not only determined by the perceived relationship, at the time of judgment, between a stimulus of the value series and a standard. They must have been affected as well by the background of perceived relationships between this particular stimulus and all other stimuli of the same series. This assumption is further supported by the evidence that the effects exercised by a particular stimulus on judgments pertaining to a series of stimuli increase as a function of the extent to which this stimulus is perceived as forming part of the series which is being judged (**4**).

In other words, in the "relevant" experiments on overestimation two aspects of the situation must be taken into account: the *inter*serial and the *intra*serial. The first consists of the perceived relationships of magnitude between any stimulus of the value series and the neutral stimuli; the second is concerned with the perception of relationships between the stimuli of a value series as compared with the corresponding relationships in an objectively identical neutral series.

The phenomenon of accentuation of differences between the stimuli in a value series can be isolated in its "pure" form when the following two requirements are satisfied: the subjects' judgments must be based on comparisons between the

various stimuli of this series, and not between these stimuli and some extraneous ones; and the value and neutral series must be objectively identical, value being the only experimental variable in which they differ. These requirements were satisfied in some experiments recently completed by the writer (27).[3]

An ordered series of ten weights was used, and the subjects were requested to judge their heaviness in terms of seven category numbers. Each subject underwent an equal number of sessions under the "value" and under the "neutral" conditions. An experimental session consisted of two parts: in the first part, the entire series was presented several times in random order, but no judgments were reported by the subject. In the second part, following the first after an interval of about three minutes, all the stimuli were presented again several times, and judgments of weight were requested at each presentation. In the value condition, a small paper bonus (gift certificate, exchangeable for a book) accompanied each presentation of one of the two heaviest (or two lightest) stimuli of the series during the first part of the session. No rewards were given in the first part of the neutral sessions. In this way, the effects of the two experimental conditions, introduced in the first part of the sessions, on judgments of weight in the second part could be assessed. A total of 60 adult subjects were used in the four experiments. They were told, as part of the instructions, that the purpose of the experiment was to investigate the effects of monotony on the speed of performing a simple task which consisted of discriminating between weights. The "small paper tokens," which they were to receive from time to time, were being introduced in order "to vary the degree of monotony." A questionnaire, which was presented to the subjects after completion of all the experiments, revealed that none of them doubted the truth of these statements. The subsequent disclosure of the experimenter's bad faith caused considerable surprise.

Data consisted of differences in the extensions of the scales of judgment between the two conditions for each of the subjects. Results can be briefly summarized as follows: in the first two experiments, in which rewards were associated with the two heaviest and the two lightest stimuli of the series, respectively, extensions of the scales of judgments were significantly larger for the value than for the neutral condition. In a third experiment, in which rewards were associated indiscriminately with all the stimuli of the series, no such effect occurred. In a fourth experiment, the procedure was the same as in the first and second experiments, apart from the fact that the paper "bonuses," passed by the experimenter to the subjects at each presentation of either one of the two heaviest or one of the two lightest stimuli, were devoid of all value. Once again, no significant effect on the extension of the scale of judgments was observed.

All this evidence is, to say the least, strongly suggestive. The advantages of using accentuation of differences as an explanatory device for overestimation are threefold:

a. It accounts for some seemingly contradictory results, such as the underestimation of the small end of the value series reported in some studies (7, 8); it also accounts for Bruner and Rodrigues' (7) "relative overestimation."

b. It does not require an introduction of *deus ex machina* principles to account for the phenomenon of overestimation in the "relevant" class.

c. With some extensions, it can be ap-

[3] This investigation was supported by a grant from the Durham Colleges Research Fund.

plied to the "interserial" differences in the "relevant" experiments, and to most experiments on overestimation which fall outside the "relevant" class.

The first point above does not require further elaboration. The second can be supported by some lines of argument, independent of each other.

McCurdy (22) has recently pointed out that by exaggerating the differences between the various coins, memory commits a "good error." In his ingenious attempt to relate the results of the studies on coins to the concept of schemata, he limited his discussion to series of coins where strong schemata could have developed through long familiarity with the stimuli. However, this would not account for the results of those experiments in which accentuation of differences in the value series, as compared with the neutral series, was found to exist (9, 27), in spite of the fact that the conditions for the development of the schemata were, as far as one can judge, not notably different for the two series.

What is the nature of the experience with coins, or with any series of stimuli, where it is important to discriminate sharply between the elements of the series? Discriminative responses to stimuli in such series are not usually made in terms of precise quantitative labels attached to individual stimuli. They are made in terms of "larger than," or "smaller than," the neighboring elements of the series. Minimizing the differences entails a risk of confusion; accentuating them is an additional guarantee of a successful response. However, the normal routine of responding in terms of "larger" or "smaller" is upset in most experiments on overestimation. The usual technique for correct handling of coins, for example, which is based primarily on an awareness of the relevant differences between a particular coin and other coins of the series, present or absent, is not quite adequate, as unusually precise individual quantitative labels are requested in the experiments. These absolute labels, whether obtained through matching or through some kind of verbal categorizing, may be expected to reveal an accentuation of differences between the stimuli, since they would reflect, and possibly exaggerate, what is otherwise implicit in the relative judgments of comparison pertaining to the series.

It is in this sense only that the results yielded by the studies on overestimation can be treated as experimental artifacts. In this context, it is not particularly important to find out whether the stimuli are really "seen" as larger or smaller. They are *reproduced* as such; to ask the subject to match a variable standard to a stimulus, or to assign to the stimulus a quantitative verbal label, is essentially asking him to reproduce its size. This reproduction, which involves an activity very different from stating vaguely that an object is sizably larger or smaller than something else, lends itself easily to a sharpening of the relevant distinctive feature of it, which is, in this instance, its difference in size from the next object in the series. The phenomenon is not unfamiliar: Gibson reported some time ago (12) that a sharpening of differences occurred in the early stages of aircraft recognition training, when his subjects were asked to draw the silhouettes of the various aircraft.

A prediction is possible here to the effect that shifts in the estimates of magnitudes would either not occur or promptly disappear if the training to discriminate between the elements of a series, along which discrimination is important for the subject, was directed towards the accuracy of individual quantitative labels rather than towards a clear distinction in relative terms be-

tween the stimuli. This is exactly what happened in an experiment by Smith, Parker, and Robinson (25) in which accuracy of report was a condition of obtaining the reward. Prizes were offered to the subjects who would report correctly the greatest number of *dots* forming clusters which were flashed successively on a screen. In a control group, prizes went to those subjects who reported correctly on the number of dots in the greatest number of *clusters*. No information about the accuracy of their estimates was available to the subjects during the experiment. The first group showed overestimation in the early stages of the experiment, but after a certain number of trials the performance of both groups converged.

An experiment is being designed at present to test a further implication of the present argument: that in any ordered series, training to discriminate in comparative terms between the stimuli will result in larger estimated differences between the elements of the series than training to label the stimuli quantitatively in familiar units of measurement.

These assumptions of a "functional" mechanism underlying the accentuation of differences may be supplemented by some evidence coming from a different quarter. It consists of the findings about the effects of multidimensionality on the acuity of discrimination along a series. Eriksen and Hake (11) have reported recently that when the method of absolute judgments is used, the number of discriminable steps for a series of stimuli is greater when the stimuli vary concurrently in two or more dimensions than when they vary in one dimension only. More specifically, they found that discriminability was considerably greater when the "stimuli varied in size and hue, size and brightness, hue and brightness, and size, hue, and brightness" (p. 159) than when they varied in only one of these dimensions. Eriksen and Hake suggest that their results were due to a kind of summation, to "the ability of *S*s to make fairly independent judgments of stimulus values along each of the component dimensions" (p. 158). They add, however:

> We cannot assume that this is always the case for compound stimuli. For some stimuli, judgments made of values in the separate dimensions may be interrelated. That is, the evocation of a particular response tendency by a component of a compound stimulus may change the likelihood of evocation of other response tendencies by other components of the stimulus. We could expect this to occur when *S*s have learned by long experience to associate the occurrence of certain values in one dimension with the occurrence of particular values in another.

This interrelation may account in part for such perceptual "errors" as the size-weight illusion. Ryan (23, quoted by Hochberg, 15) writes: " . . . *O* is not really reporting *weight* at all, but *specific gravity*, and the illusion is really not an illusion at all. *O* is simply unable to report on weight being influenced by the striking differences in density involved." In the field of size estimation, Holzman and Klein (16) have recently reported that heavy discs were judged larger by their subjects than light discs of identical size, and that black discs were judged smaller than grey discs of identical size and weight. They add wistfully that these results " . . . do not exclude the possibility that value and need may indeed be important variables in size judgments" (p. 40).

Two conclusions can be drawn from the above: the first, supported by Eriksen and Hake's results, is that under some conditions of judgment, and in a situation new to the subject, compounding of concurrently varying dimensions will result in a clearer perception of differences between the stimuli along a series. The second is related to the first, and already outlined in the above quotation from Eriksen and Hake: when

the association between dimensions has been a long-standing one, new training may not be capable of inducing changes in discriminability. In such cases, judgments along one of the dimensions have already been influenced, previous to the experiment, by concomitant changes in other dimensions. In other words, it is likely that the Eriksen and Hake situation represents the incipient stages of a process which, if given an opportunity to develop further, might lead to interrelations such as those that Ryan suspects to be responsible for the size-weight illusion, or that Holzman and Klein (16) found between the estimated size and the actual weight of their discs. Interrelations of this kind exist, of course, when there is a consistent pairing, in the environment, of changes in more than one dimension.

Such pairing exists by definition in series where magnitude and value vary concurrently. The difficulty is that value is not a "dimension" in the physical sense. It is, however, an important attribute of the stimuli in such series, if only because efficient discrimination between the stimuli in terms of differences in value is, in most cases, more important than discrimination in terms of the physical dimension. It is not reasonable to expect that subjects who, in an experiment, judge a series of coins in terms of size, judge it in terms of size alone. This would be a feat of abstract behavior difficult for adults, and even more so for children. As Vernon (28) points out in her discussion of the experiment by Bruner and Goodman (5): "It seems fairly certain that . . . there was an inability to isolate a single aspect from a global percept" (p. 187). Or, as Hochberg writes in another context (15): " . . . the subjects are responding in terms of a dimension which specifies the most about the stimulus series with the least number of categories, rather than in terms of the dimension expected by the experimenter."

In the experiment on weights described earlier (27), value is the only "dimension of difference" between the two series. It would also be a supplementary and important difference between two series which differ not only with regard to value, but also with regard to some other dimensions. In the first case, the difference in value is the only contributing factor to a more pronounced accentuation of differences in the value than in the neutral series; in the second, it is one of the contributing factors. Once again, this does not necessarily mean that the subjects "see" the stimuli in the value series as being more different from each other, or larger, or smaller. They respond *as if* they perceived them as such, as this is as far as the phenomenology of it can go for the time being.

"IRRELEVANT" DIMENSIONS AND THE ACCENTUATION OF DIFFERENCES

The accentuation of differences *between* the stimuli of a value series may partly account for the results obtained in the "relevant" experiments on overestimation. However, it has no direct bearing on the problems raised by the results of the "irrelevant" group of experiments; nor does it help to understand why in the "relevant" group a rather consistent trend towards overestimation of the value series should be superimposed, as it were, on the sharpening of differences within it.

The preceding discussion is not, however, entirely irrelevant to these problems. As was pointed out earlier, in a "relevant" series the subjects' judgments are simultaneously determined by the perceived differences between the stimuli of the series, and between a particular stimulus of the series and an extraneous stimulus. In the "irrelevant"

experiments, where the magnitude under investigation seems in no way related to the presence or absence of value, the fact that one stimulus is "valued" and another "neutral" would form an additional distinctive feature between them, superimposed on the physical differences. Examples of such experiments have been given above (6, 19, 20). Beams' study (2) on overestimation of size of favorite food objects by children provides another example. The common feature of these experiments is that the subjects are requested to judge along a dimension which, in usual circumstances, does not help to discriminate between the objects in terms of their value. Children, who in the experiment by Lambert *et al.* (20) were asked to report on the size of the chips, had just experienced a situation in which the value of the chips was determined by their color. The refugees, who in the study by Klein *et al.* (19) were judging the size of discs containing the swastika, would be unlikely, in any other circumstances, to pay special attention to the *size* of the emblem.

In these experiments the valued and the neutral objects belong to sharply distinctive categories of experience, and size is certainly not the basis of classification. In such cases, to judge size exclusively on the basis of size would, once again, be possible only if the subjects were able to abstract one element, and not an important one, out of a complex compound of experience. A more likely assumption is that the qualitative "motivational" differences between the two kinds of stimuli function in the direction of accentuating the perceived differences between their "irrelevant" magnitudes.

This would lead to the prediction that, all else constant, the perceived differences of magnitude between valued (or otherwise relevant) and neutral stimuli would be larger than the corresponding differences between two sets of neutral stimuli. It does not, however, allow for the prediction that the valued stimuli would be perceived as *larger* than the neutral ones.

Some interesting indications which may suggest a way out of this difficulty can be found in the experimental literature. In the group of studies on "irrelevant" magnitudes, three experiments seem typical of those which yielded negative (or partly negative) results. Klein *et al.* (19) reported that their subjects did not reliably overestimate the discs containing the swastika. Bevan and Bevan (3), who used a "quasi-representative design" and a heterogeneous assortment of objects with their two children, found that the ratio of estimated to real size was not higher for the liked than for the disliked or neutral objects. Finally, Lysak and Gilchrist (21) found that increasing complexity of design was a significant determinant of increasing overestimation; but dollar bills were not judged larger than the corresponding rectangles, nor was there any correlation between overestimation and value of various dollar bills. Experiments by Lambert *et al.* (20) and by Beams (2), already referred to, are representative of those in which positive results were reported.

The experiments of Klein *et al.* (or at least the parts of them involving the swastika) and of Bevan and Bevan (3) have one feature in common: the value of the objects used has nothing to do with their size. The data, as they are presented, do not allow one to draw conclusions about an accentuation of differences between the sizes of valued and neutral stimuli; but if there was such an accentuation, there are no grounds to assume on the basis of the present argument that it should work, with any degree of consistency, in the direction of overestimating the valued objects.

On the other hand, it seems reasonable

to say that the dictum "the bigger the better" expresses a profound truth to Beams' 10- to 12-year-old *Ss* when they are confronted with a piece of fruitcake, a sugar cookie, or a "marshmallow-covered chocolate cupcake" (2, p. 197). Lambert *et al.* worked with much younger children (3 to 5 years old); the number of chips of a particular color which they received was related to the amount of candy which was the final compensation for their efforts. It would not be surprising if this, in view of the age of the subjects, had a great deal to do with their judgments of size.

The negative results of Lysak and Gilchrist (21) lend further support to the thesis that the accentuation of differences will tend consistently towards an overestimation of valued objects *only* when there exists a valued series in which there is some discernible relationship between magnitude and value. Their adult subjects knew from long experience that the value of dollar bills does not vary as a function of their size. They may have accentuated the differences in size between the bills and the corresponding rectangles, but not necessarily or consistently by overestimating the bills more than the control stimuli.

A comparison of this study with other experiments performed on adult subjects helps to bring out more clearly its particular features which may account for the lack of positive results. In the experiment by Klein *et al.*, there is no tendency to overestimate the disc bearing the swastika; but in one of their two experimental situations, all results pertaining to the disc bearing the dollar sign are significant (19, Table 3, p. 103). This is further substantiated by Solley and Lee (26) who attempted " . . . to determine whether the differences in perceived size of discs with symbols drawn upon them, as reported by Bruner and Postman, could be better explained by their hypothesis of symbolic value than by the more general Gestalt principle of closure" (p. 142). They report an overestimation of the discs containing a dollar sign (as compared with discs containing a figure of presumably equal degree of closure), but no such effect with the discs containing a swastika. The fact that a symbol of money drawn on a disc is strongly suggestive of the familiar series of coins, whose value increases with their size, may well be responsible for these results.

The negative outcome of the experiment by Lysak and Gilchrist must also be taken in conjunction with an almost identical experiment by Dukes and Bevan (10), who reported positive results. They used a number of rectangular cards, identical in size but differing in value, and engaged in "gambling" with their subjects. A number representing its value was stamped on each of the cards. The cards were drawn at random from a bag, and at each draw the subject won or lost an amount of money proportional to the positive or negative number printed on the card. As soon as each drawn card was returned to the bag, the subject was asked to match its size with one of a series of blank "reference" cards spread out in front of him. Dukes and Bevan's hypothesis was that "as the monetary value printed on the test card increases from zero, the subject, regardless of winning or losing, will tend to select larger reference cards as his estimates" (10, p. 47). This was confirmed: a significant correlation emerged between the ratios of estimated to actual sizes and the monetary value, positive or negative, of the test cards.

The situation was almost identical with one of the arrangements of Lysak and Gilchrist's experiment (21). If the fact of conflicting results were to be ascribed, here or in other cases, to relatively unimportant differences in procedure, attempts to generalize results from a great number of psychological

experiments would have to suffer a serious setback. There is, however, one essential difference between the two experiments: Lysak and Gilchrist's subjects *knew* that all the dollar bills, with which they were presented, were of equal size. This bit of information about the cards used was not available to Dukes and Bevan's subjects. What is more, Dukes and Bevan included in their instructions the following sentence: "Here is a bag containing small white cards, *varying slightly in size*" (10, p. 44, italics mine). It would be only natural for the subjects to assume that this variation followed some regular pattern, and that this pattern consisted of an increase in size corresponding to the increase in value. Here again, a continuous variation along the dimension of value has carried over to the judgments of the physical dimension, and this has been further helped by a hint from the experimenter.

But there is some evidence that the same could have happened even without the hint. In an experiment by Ashley, Harper, and Runyon (1), a "grayish metal slug" was presented several times to the subjects. At different presentations of the slug they were told that "it was made of either lead, silver, white gold, or platinum" (p. 568). Judgments of the size of the slug reflected its imaginary value: the size of the variable standard "that was called equal to the slug increased as the cost of the metal increased" (p. 572). As in the preceding example, a progressive increase in value is related to an imaginary increase in size, while the actual size is kept constant.

The evidence summarized so far lends strong support to the view that both the positive and the negative results of the studies on overestimation can be best understood in terms of a hypothesis of accentuation of differences. This accentuation would work along two axes: between the stimuli of a series in which there is a concomitant variation of value and magnitude; and between the valued stimuli and the neutral ones. In the latter case, the accentuation would tend towards a relative overestimation of the valued stimuli only under certain conditions, which have been discussed above.

SOCIAL PERCEPTION

Some implications of the present discussion go beyond the problems raised by the phenomenon of overestimation. It may be said that, in a sense, "overestimation" as discussed in this paper is a special case and a convincing experimental paradigm of a more general aspect of social perception. Many social objects and events are sharply classified in terms of their value or relevance. When judgments concerning some quantifiable or ratable aspects of stimuli which fall into distinct categories are called for, differences in value or relevance cannot fail to influence the quantitative judgments in the direction of sharpening the objectively existing differences between the stimuli. A very similar conception has recently been formulated by Hochberg (15), and applied by him to the perception of in- and out-group individuals:

> If a group of individuals is perceived as different from the non-group individuals, the perceived differences between those within the group and those outside the group will automatically be sharpened, and the differences perceived between the members of the group (i.e., intragroup differences), and between those outside the group will be lessened.

These judgmental effects of categorization are probably fairly general; it is likely, however, that they are particularly pronounced when judgments are made in dimensions in which scaling in magnitude is simultaneously a scaling in value. Thus, it may well be that an accentuation of differences in size will

hardly occur between two paintings, one liked and one indifferent or disliked. But when skin color, or height, or some facial traits of social "value" are concerned, there will be a marked sharpening of differences in the degree of these characteristics perceived as belonging to individuals who are assigned to different categories. Some evidence of this is provided by a recent study on "perceptual accentuation and the Negro stereotype" conducted by Secord, Bevan, and Katz (24). Their results suggest that a group of prejudiced *Ss* sharpened the differences in the degree of Negroid physiognomic traits possessed respectively by Negroes and whites, more than did a group of nonprejudiced *Ss*. It is likely that the same is happening in the case of more abstract social judgments which are implicitly quantitative, such as, for example, those concerning the relative frequency of crimes in various social groups, as perceived by people who have an axe to grind.

SUMMARY

The thesis put forward in this paper is that the apparently conflicting results of studies concerned with perceptual overestimation can be best understood if this phenomenon is considered as a special case of accentuation of perceived differences. This would work in two directions. First, in a series of stimuli in which there exists a concomitant variation of value and of some physical dimension, the perceived differences between the elements of the series would be larger than in a corresponding neutral series. Secondly, the perceived differences in magnitude between the stimuli which possess the attribute of value and those which do not would tend to be more accentuated than the corresponding differences between neutral stimuli. This would not, however, always result in overestimation. The sharpening of differences resulting in overestimation of valued stimuli would occur only when, in the class of stimuli with the attribute of value, an increase in some physical magnitude is correlated with the increase in value. In other cases, there are no grounds for assuming that an accentuation of perceived differences between the two classes of stimuli would tend consistently towards an overestimation of the valued stimuli.

In this way, overestimation is seen as one instance of a more general phenomenon. The accentuation of differences between classes of stimuli occurs when these stimuli differ in some respects other than the dimension along which the subjects are reporting their judgments of quantity. The presence or absence of "value" or "relevance" is one such contrast. It is probable that this contrast is responsible not only for the results obtained in the field of overestimation. It may also lead to the sharpening of differences between the quantifiable or ratable aspects of social objects and events. This occurs when the dimension along which judgments are made is not the primary basis for assigning the various stimuli into their sharply distinctive categories.

REFERENCES

1. ASHLEY, W. R., HARPER, R. S., & RUNYON, D. L. The perceived size of coins in normal and hypnotically induced economic states. *Amer. J. Psychol.*, 1951, **64**, 564–572.
2. BEAMS, H. L. Affectivity as a factor in the apparent size of pictured food objects. *J. exp. Psychol.*, 1954, **47**, 197–200.
3. BEVAN, W., & BEVAN, D. C. Judged size and personal relevance: an exercise in quasi-representative design. *J. gen. Psychol.*, 1956, **54**, 203–207.
4. BROWN, D. R. Stimulus-similarity and the anchoring of subjective scales. *Amer. J. Psychol.*, 1953, **66**, 199–214.
5. BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in per-

ception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.

6. BRUNER, J. S., & POSTMAN, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203–208.
7. BRUNER, J. S., & RODRIGUES, J. S. Some determinants of apparent size. *J. abnorm. soc. Psychol.*, 1953, **48**, 17–24.
8. CARTER, L. F., & SCHOOLER, K. Value, need, and other factors in perception. *Psychol. Rev.*, 1949, **56**, 200–207.
9. DUKES, W. F., & BEVAN, W. Accentuation and response variability in the perception of personally relevant objects. *J. Pers.*, 1952, **20**, 457–465.
10. DUKES, W. F., & BEVAN, W. Size estimation and monetary value: a correlation. *J. Psychol.*, 1952, **34**, 43–53.
11. ERIKSEN, C. W., & HAKE, H. W. Multidimensional stimulus differences and accuracy of discrimination. *J. exp. Psychol.*, 1955, **50**, 153–160.
12. GIBSON, J. J. Social perception and the psychology of perceptual learning. In M. Sherif & M. O. Wilson (Eds.), *Group relations at the crossroads*. New York: Harper, 1953. Pp. 120–138.
13. GILCHRIST, J. C., & NESBERG, L. S. Need and perceptual change in need-related objects. *J. exp. Psychol.*, 1952, **44**, 369–376.
14. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.
15. HOCHBERG, J. E. Psychophysics and stereotypy in social perception. In *Emerging problems in social psychology*, in press.
16. HOLZMAN, P. S., & KLEIN, G. S. Intersensory and visual field forces in size estimation. *Percept. Mot. Skills*, 1956, **1**, 37–42.
17. JENKIN, N. Affective processes in perception. *Psychol. Bull.*, 1957, **54**, 100–127.
18. JOHNSON, D. M. *The psychology of thought and judgment*. New York: Harper, 1955.
19. KLEIN, G. S., SCHLESINGER, H. J., & MEISTER, D. E. The effect of values on perception: an experimental critique. *Psychol. Rev.*, 1951, **58**, 96–112.
20. LAMBERT, W. W., SOLOMON, R. L., & WATSON, P. D. Reinforcement and extinction as factors in size estimation. *J. exp. Psychol.*, 1949, **39**, 637–641.
21. LYSAK, W., & GILCHRIST, J. C. Value, equivocality and goal availability. *J. Pers.*, 1955, **23**, 500–501. (Abstract)
22. MCCURDY, H. G. Coin perception studies and the concept of schemata. *Psychol. Rev.*, 1956, **63**, 160–168.
23. RYAN, T. A. Interrelations of the sensory systems in perception. *Psychol. Bull.*, 1940, **37**, 659–698.
24. SECORD, P. F., BEVAN, W., & KATZ, B. The Negro stereotype and perceptual accentuation. *J. abnorm. soc. Psychol.*, 1956, **53**, 78–83.
25. SMITH, K. R., PARKER, G. B., & ROBINSON, G. A. An exploratory investigation of autistic perception. *J. abnorm. soc. Psychol.*, 1951, **46**, 324–326.
26. SOLLEY, C. M., & LEE, R. Perceived size: closure versus symbolic value. *Amer. J. Psychol.*, 1955, **68**, 142–144.
27. TAJFEL, H. The role of value in the formation of a scale of judgments. *Bull. Brit. psychol. Soc.*, 1956, **29**, 14. (Abstract)
28. VERNON, M. D. The functions of schemata in perceiving. *Psychol. Rev.*, 1955, **62**, 180–192.

(Received September 21, 1956)

Psychological Review
Vol. 64, No. 3, 1957

A MECHANICAL MODEL FOR HUMAN ATTENTION AND IMMEDIATE MEMORY [1]

D. E. BROADBENT

Applied Psychology Research Unit, Cambridge, England

Many people have a natural distaste for model building. A mechanical model is essentially a theory expressed in material parts rather than in abstract symbols such as words or mathematical expressions. Its logical standing is the same: that is, it stands or falls by the degree to which it fits the results of experiments on human or other animals. Yet many models in the past have been somewhat undistinguished in the closeness with which experiment has been considered in their design. It is difficult to avoid feeling that this is because a model is unduly laborious to build as compared with a verbal theory, so that the builder tends to become obsessed with the properties of his model rather than those of the organism. Consequently there is much to be said for building theories verbally, and especially for using the qualitative terms of information theory in the hypothetico-deductive fashion ably set out by Mackay (29). Such an approach has certain advantages over the other popular alternative of quantitative S-R terms; these advantages have been considered elsewhere (11).

For example, the writer holds that the human perceptual system has a limited capacity, that in consequence a selective operation is performed upon all inputs to the system, and that this operation takes the form of selecting all inputs having some characteristic in common. Such an operation extracts little information from the signal and thus should be economical of nervous mechanism. Characteristics on which the selection can operate may be named "sensory channels." The particular selection made at any one time will depend partly on characteristics of the input itself (physical intensity, earliness in time, absence of recent inputs on that channel, position of the channel in the hierarchy of all channels) and partly on information in a more permanent store. The change from one selection to another will take a determinate time.

Incoming information may be held in a more temporary store at a stage previous to that of the selective operation. Such information will pass through the perceptual system on the next subsequent selection of the sensory channel of its arrival, if it is still in store; but the probability of the latter condition being fulfilled will decline with time spent in store. After passage through the perceptual system, information may be returned to the same temporary store, the selection of information for such return being determined by information in a more permanent store.

But although such a purely verbal theory may fit experimental results, it is difficult to communicate to others without putting them to the trouble of learning the necessary vocabulary. And if the theory is rephrased, still abstractly, it is open to misinterpretation; thus Deese (19), in a paper which makes a number of valuable contributions to the theory of prolonged work, has described the writer as postulating an inhibitory construct in human per-

[1] This work was supported by the British Medical Research Council, and the writer works under the general direction of Dr. N. H. Mackworth. He has discussed the topic with profit with many individuals; Dr. J. Brown should be especially mentioned.

formance. This misunderstanding probably arises from the fact that the theory outlined in the last two paragraphs is intended to apply both to conditioning and to human watch keeping; in neither case is it thought necessary to find an inhibitory construct of orthodox learning theory type. To say that conditioning and perception are related is not to apply a particular interpretation of the former to the latter.

Clearly, then, some sort of expository device is needed for an abstract theory using unfamiliar terms. And it is even difficult for the theorist to remember in abstract form the results of the many different experiments which a good theory should consider. A simple mechanical model has the virtue of avoiding these difficulties. It has other vices: it may have accidental properties which mislead research. Perhaps the best compromise is to state a theory in abstract terms, and also to give a model which can be described by the same verbal theory. Information concepts are applicable to any system, whatever its physical nature, and so may equally fit a model or a man. This is the approach of Deutsch (20, 21), and it has real advantages, independent of the value of his particular theory. The present paper is therefore intended to describe an extremely simple model of the human perceptual system. It may serve both as an easy introduction to the formal theory in information flow terms and also as a convenient mnemonic for the results of a number of experiments.

THE BASIC MODEL

The necessary requirements are a Y-shaped tube (Fig. 1) mounted vertically, and a set of small balls. Each ball bears a number so that all are individually recognizable. The Y tube has a narrow stem which will just take only one ball, though the branches are wider. At the junction of stem and branches is a hinged flap which normally hangs straight downward, but which can be pivoted about its upper edge so as to close off either of the branches of the Y. This pivoting can be done by a handle from outside the tube, controlled by stored information; purists may wish to control the handle by a punched-tape system, but a human being is an adequate substitute. When the handle is left alone the flap moves freely so that a ball dropped into one arm of the Y will knock the flap aside and fall into the stem of the Y.

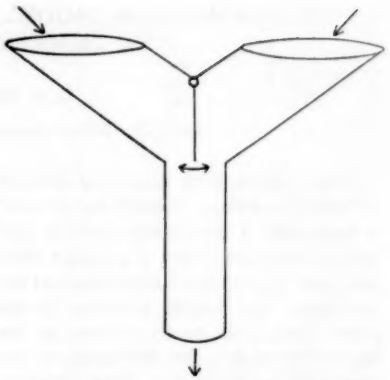


FIG. 1. The simple model for attention.

In this model the balls represent the information from various stimuli. The branching arms represent different sensory channels; thus one might be one ear and one the other ear. Alternatively, one might be the ear and one the eye. (Sensory channel is not, however, quite equivalent to sense-organ, since we would treat sounds localized in different places as being on different channels.) The bottom of the Y represents a response output, so that the process of dropping a ball into the arms and observing its emergence at the bottom is analogous to that of delivering a stimu-

lus and observing a response. The behavior of the model resembles that of man in the following ways:

(*a*) If two balls are dropped simultaneously, one into each of the branches, they will strike the flap on both sides; it will not move and therefore they will jam in the junction. Numerous experiments show this "distraction" effect, but there are certain advantages in citing an auditory experiment (**5, 33**). These same experiments show that if the handle is used to shut off one branch before the balls are inserted, then the ball entering the other branch will emerge successfully, which is analogous to previous instructions.

(*b*) If the two balls are not strictly simultaneous, the first to arrive will obtain an advantage by knocking the flap over and shutting out the other. This had been shown to be analogous with competing auditory stimuli by Spieth, Curtis, and Webster (**36**).

(*c*) If the Y is not perfectly vertical, the ball in the more vertical branch will have an advantage over a simultaneous ball in the other because the door will hang to one side. Equally, one sensory channel may have an advantage over another, as has been shown for high-pitched noise as compared with low (**10**).

(*d*) If one ball is flung violently down its branch, it may succeed in forcing over the door against the unassisted weight of a ball on the opposite branch. Equally, an intense stimulus may have an advantage (**2, 10**).

(*e*) After a single ball has been passed through the system, the door will swing back from the position into which it has been pushed. Naturally it will overswing, and temporarily close the branch which has just been used. A stimulus has similarly an extra advantage for response if it comes on a previously quiet channel as opposed to a previously busy one. This has been shown by Poulton (**34**) for auditory signals, and a related finding is that of Hyman (**25**) for visual reaction times. In the latter case, stimuli of different frequencies of occurrence were delivered, and it was found that average reaction time to a set of stimuli was proportional to the information conveyed by them; but the infrequent signals, while giving long reaction times, did not give times as long as those to be expected from information theory calculations. Note, however, that the time taken for a swing is important in the model; this will probably also be true in man, to judge from data which are best considered below. It may also be connected with Hyman's finding that, for numbers of alternatives greater than two, the second of two identical signals received an unduly fast reaction.

(*f*) If a given number of balls are to be put through the tube, it is best to deliver them asymmetrically, the majority to one branch. There is then less risk of jamming than if equal numbers are admitted by both branches. The analogous finding for auditory messages has been reported by Webster and Thompson (**37**). This point is related to another, that if balls are being inserted into one branch at random intervals the effect of increasing the rate of delivery of balls through that same branch (the "speed" of work) is not the same as that of adding the same number of extra balls to the other branch (the "load" of work). There is more risk of jamming in the latter case. This point was first clearly emphasized by Conrad, using visual signals (**16**). Obviously the effect of using two branches rather than one will be more serious if the rate of delivery of balls is high, since this increases the probability of a jam. Conrad showed such an interaction of speed and load. Mackworth and Mackworth (**28**) have shown a similar

effect in a different type of task, and have demonstrated that the fluctuating difficulty of this complex task at any instant can be represented by the amount by which each signal is overlapped by other simultaneous ones.

At this point we may pause to consider the cynical reader who is wicked enough to be doubtful of the existence of a Y-shaped tube somewhere in the region of the thalamus. Such a reader will probably have noticed already a situation which will follow when two balls are dropped into the model and jamming is avoided by the door being to one side for one of the reasons listed. The favored ball will descend the stem of the Y, but the impeded ball will not therefore disappear. It will emerge later, when the door next swings back to the opposite branch. Surprisingly enough, this also happens with man. Simultaneous stimuli either jam or produce successive responses.[2] It has been shown for three different combinations of sensory channels (7, 8), and is probably an effect identical with the "prior entry" of classical psychology. In the same way Conrad has not only shown complete failures to respond (17) but also shifts in the time at which responses appear (18).

Having thus demolished the cynic, we must add that a slight complication should be added to the model in order to cover perfectly the results of experiments on successive response to simultaneous stimuli. But this will be left until the section on immediate memory; let us first deal with the effects of speed and of prolonged performance on the simple model.

(*g*) As the stem of the Y is so narrow, there will be a certain amount of delay between the insertion and emergence of each ball. At slow rates of insertion each ball will, however, emerge before the next is inserted. At faster rates there will develop a lag such that one ball may be inserted before the previous one emerges. With still faster rates balls will begin to accumulate in the branch; response will get further and further behind stimulus as the task proceeds. But this is a desperate expedient and will lead to breakdown when the branch is full unless the rate slows down again. These stages have been shown by Vince (35) for visual stimuli; a related effect has also been shown in less detail in hearing (4). A result of the piling up of balls in the branch when the rate is too high is that a ball inserted at a very short interval after another will remain in the tube longer than is normal: the "psychological refractory period" (38). It must be remembered that balls are analogous to information and not to stimulation; highly probable stimuli will not give an unduly long reaction under these conditions (22).

(*h*) Now suppose that we wish to operate the handle in such a way that one branch always has priority: prolonged performance of a task involving only one sensory channel. If there is a fairly rapid flow of balls down the selected branch, the handle will hardly need holding after the initial setting, since the beginning of any swing back will probably be checked by another ball. But if the selected branch is not very busy, the weight of the door must be held by a positive force on the handle. In this case the fingers holding the handle will fatigue; for purists, the

[2] At the risk of complicating the issue, we must say that "stimuli" in this sentence means "stimuli not of low information content." A familiar predictable sequence of stimuli may quite well produce responses simultaneous with other responses (1). And the same S-R unit interferes less with another task when drawn from a smaller ensemble of possibilities (12). This point is of course implied in the statement that the perceptual mechanism has a limited capacity in the sense of information theory.

punched-tape machine will only hold one branch shut for a limited period. When the handle is released the door will swing back to close the opposite branch, and then return pendulum-wise to the desired position when it will again be held. The result will be that prolonged tasks in which unexpected stimuli appear for less than a certain period (the "swings of the door") will show marked decrements in performance. But similar tasks in which the stimuli are present for longer times will show much less decrement.

The earlier evidence for this view of prolonged performance has been summarized elsewhere (6). But earlier presentations have misled Deese (19) into supposing that this view implied a decrement with continued performance of *any* vigilance task. He employed tasks in which a signal was either presented repeatedly until seen, or else painted on a tube face by a sweep line which did not return to obliterate the signal for nearly three seconds. It was in fact predicted by the writer (6, p. 300) that such tasks would show little or no decrement, and this was found by Deese to be the case. The latter type of task gave more signs of decrement, probably being a borderline case, since the trace left on the phosphor screen decayed until scarcely visible at the end of the three seconds. It should be noted that an expectancy theory of the type favored by Deese, while undoubtedly applicable to some aspects of vigilance, is not able to account for the effects of varying the length of signal presentation. In our Y-tube model, the role of expectancy is incorporated by describing the balls as representing information rather than stimulation; the more probable signals receive more efficient response. But the swinging of the flap is also necessary.

To summarize the writer's present views on vigilance, the efficiency of a man asked to detect infrequent signals should be described by both a mean and a variance. The mean level is determined by such factors as the rate at which signals arrive but not necessarily by the length of time since the session began. The variance, on the other hand, increases as the session progresses, short intervals of extremely low efficiency being interspersed with fairly long periods of normal or supernormal efficiency. The score from any given task may depend on one or on the other of these quantities. Thus, for example, the earlier British work has been mostly concerned with the instants of very low efficiency; the type of results reported by Deese and by later British workers (6) mostly with the mean over appreciable periods of time.

To return to our model, two further points should be made. The first is that the door need not swing the whole way over when the handle is released; it may reach a central position just as two balls arrive. There will then be a jam rather than passage of the wrong ball. Equally, failure of reaction to a task need not imply overt reaction to some irrelevant stimulus.

Secondly, as the balls represent information rather than stimulation, tasks in which the sequence of stimuli is predictable will not show fatigue decrements of this type. To show decrements, the signals must be unpredictable either in content, as were those of Bills (3), whose "blocks" are instances of this effect, or else in time of occurrence, as were those of Mackworth (27); this, however, is only one of the necessary conditions.

IMMEDIATE MEMORY

Our Y tube does seem at this stage to have related a number of facts about perception and put them in a way which most people can understand. It is admittedly ludicrous as a description of

what really happens in the brain, but this is a positive advantage. Psychologists are not likely to mistake this model for speculative neurology, and so they should concentrate their experiments on the essentials of the theory rather than the irrelevant properties of the model. As a device for communicating the outline of the theory, however, the model seems sufficiently adequate to justify an extra complication in order to express a theory of immediate memory. This theory is in a slightly different position from the views on perception which have been given so far; the latter are entailed by the experiments, but the theory of immediate memory is not the only conceivable explanation of the observed facts. Yet it has a fairly high probability and is worth discussing.

The complication is twofold. First, a device must be supposed fitted to each branch of the Y tube, such that if any individual ball remains in the branch for more than a certain time continuously it is removed from the system completely. This could be done by filling the tube with acid, but the writer does not wish to encourage the development of a race of fingerless psychologists. Mechanical devices are quite possible, though complicated to describe, and the details will therefore be left unspecified and available from the author. The second complication is that from the foot of the stem two return tubes lead back to the branches (Fig. 2). Admission to these return tubes is controlled by a lower door which again is operated by a handle. The latter is dependent on an outside operator, or the familiar punched-tape machine—in brief, some form of stored information. Finally, as the return tubes are operating against gravity they must contain some form of conveyor, but this has no particular psychological significance.

In immediate memory experiments, as usually performed, a stream of stimuli

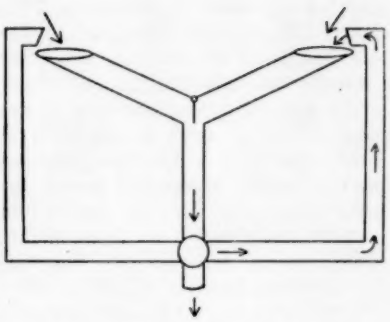


FIG. 2. The model modified to illustrate the theory of immediate memory as a recurrent circuit; or, in other terms, as a fading trace periodically revived by rehearsal.

is delivered completely before any overt response is required. Thus all the information in the stream is within the organism at one time; all the balls are somewhere in the tubes. Yet with small numbers of balls the time between insertion and emergence can be varied without any apparent effect on efficiency (13). Let us consider some of the methods which may be adopted.

(*a*) In the first place, let us suppose a short series of balls to be waiting in one of the branches. If they stay there indefinitely, the critical time will be exceeded and the ball which has been there longest will leave the system. Thus the branch will not serve as a store during an unlimited period of delay before response is allowed. What is perfectly possible, however, is for the series of balls to pass down the stem and, by the lower door, back up a return tube to the other branch. As soon as the full series has completed this round the process can be repeated, so that the set of balls can be kept in the system for any desired time. But note that insertion of any other ball during that time will either mean that the extra ball will never emerge, or else that the circulation of the series will meet inter-

ference. Note also that the system will only operate with a limited number of balls; above that number each ball will have to wait in the branch for more than the critical time, and there will therefore be a severe breakdown of storage. These are well-known characteristics of immediate memory. Less obvious is the suggestion that the interfering effect of an extra ball during the delay will depend on the size of the series circulating, being most serious near the limit. This has been shown by Brown (**14**). He also demonstrated that the interval before or after the interpolated stimuli had little effect, as one would expect, and that if the extra stimuli came before the memory span stimulus there was little effect. It is Brown's explanation of his results which has prompted the present account.

(*b*) If both branches contain balls, movements of the door will put one group into circulation before the other; to move the door over while a group is actually passing would be to risk a jam. So all balls on one branch will pass down the stem before any on the other. As has been said, the memory analogy is in fact true, if a digit memory span is obtained with half the digits on one ear and half on the other, or half on the eye and half the ear (**7**, **8**). In addition, the first branch to be dealt with undergoes less risk of any ball reaching the critical time; equally, performance is better on the first set of digits to be recalled (**9**, **13**).

(*c*) So far we have considered only the case in which the branches start with balls in them. But in practice these balls will be inserted into the branch one after another, and the extra time intervals thus produced will have their effect. Thus, for example, if balls are inserted rapidly into both branches simultaneously, the difficulty of moving the door will apply, as mentioned in the last paragraph; and so all the balls in one branch will be dealt with before any in the other. But if there is a sizeable interval between successive balls in each branch, the door may be swung back and forth so as to deal with each branch alternately. A slow presentation of stimuli may equally allow them to be dealt with in the actual order of arrival rather than channel by channel (**7**, **8**). It should be noted that the time allotted to "swings of the door," as measured by these experiments, agrees with that determined by the experiments on prolonged performance mentioned previously (**6**).

(*d*) Again, suppose a lengthy series of balls is inserted through one branch, the door being held open for them by the handle. If a couple of balls are meanwhile delivered to the other branch and wait there till the stem is clear for them to pass, there is naturally some risk that this extra pair of balls will exceed the critical time. The risk will be greater if they are inserted with the earlier balls of the long series than if inserted with the later balls. This also is true of immediate memory for spoken digits arriving at the two ears (**9**). A point of interest concerns the effect of "irrelevant" balls which arrive on the second branch during a long series on the first. If all the balls in the second branch are irrelevant, the door may be kept closed against them and they will eventually exceed the critical time and be removed (**15**). But if some relevant balls are also on the second branch, the door will have to be opened and the mixture of relevant and irrelevant balls passed down the stem. The former can be recirculated in the usual immediate memory fashion, and the latter removed by the lower door. The presence of irrelevant balls either before or after relevant ones will therefore produce a greater risk of jamming or of the critical time being exceeded. This also is true of immediate memory (**9**).

(*e*) Now consider cases in which the order of the balls emerging at the bottom of the stem is different from that of their insertion. The extreme case is that of "backward memory span." With our simple model, the easiest situation to consider is that in which the first half of a series of balls is to emerge after the second half. This can be arranged on the first circulation by passing the first balls down the stem and back up the return tube to the branch opposite to that in which they were inserted. The later balls are returned to the same branch which they entered, and on the second circulation are the first to be admitted to the stem. After they have passed the flap the other balls can follow them. This will mean that the first balls have stayed in the branch longer than will normally be necessary for simple recirculation, and so are more likely to exceed the critical time. Therefore such a rearrangement of order will reduce the "memory span." Furthermore it will alter the order of difficulty, since the first members of the series will suffer more than the later ones. These effects have been shown by Kay and Poulton (26), and are supported by Brown (13). The reduced memory span for a rearranged list is of course familiar from intelligence testing, but the change in the serial-order effect is a more important deduction.

(*f*) Serial-order effects will clearly depend very considerably on the rates of presentation of stimuli and of required response, as compared to the rate of recirculation. They will also depend on the way in which stored information is used to operate the two handles—that is, on the strategy of the subject. They will certainly not be completely determined by the primitive learning process, as is supposed by some existing theories. In fact it has been shown by Kay and Poulton (26) that the serial-order effect is altered by the subject's absence of knowledge about the order in which recall will be required, even though the actual order remains the same as in a control experiment. (Knowledge of the amount to be recalled will also affect the efficiency of recall as well as the order, as has been shown by Brown [13]. This is because the uninformed subject must recirculate material which is not in fact to be recalled later; retention is an active process.)

Some general points about serial-order effect may be made, however. First, if the rate of response is below that of recirculation, but not so slow as to allow complete recirculation between responses, the first items in response will have stayed in store less long than the later items. So, as in (*b*) above, the earlier part of a list recalled in the order of presentation will be better recalled than the later part. Second, if the time taken to respond is eliminated —for instance, by requiring recall of only one item in the series—then the time taken to present the stimuli will be the chief factor influencing serial-order effects. The last stimuli to arrive will then be those which have been stored for the shortest time, and will be best recalled. This is the result of Gibson and Raffel (23).

Third, a special case of some importance arises when fresh material is being presented while earlier material is being recalled. Considering the model, suppose a short series of balls are inserted and recirculated to the branch opposite that by which they arrived. If they are now to be passed down the stem during the arrival of fresh balls on the original branch, the earliest of the second series of balls will suffer the longest delay. Consequently the end of the first series and the beginning of the second series will be the points of greatest difficulty. If the two series are considered as one long one, the familiar U-shaped serial-order effect will appear.

This has been demonstrated by Poulton (32). In conventional learning experiments subjects are not instructed to rehearse (recirculate) the earlier items during presentation of the later ones. But it seems plausible that, during the presentation of a long series at a medium rate, the presence of the earlier balls in the second branch should normally encourage a tendency to recirculate them while the later balls are still arriving. This would be comparable to the higher priority of a previously quiet channel, which was mentioned in the section on perception. It would result in the U-shaped curve of difficulty being the usual one for serially presented material; but the curve will be subject to great modification by instructions. Furthermore a very slow rate of presentation, resulting from allowing recirculation during gaps between presentations, will minimize the effect. This is the case (24).

LIMITATIONS OF THE MODEL

Certain properties of the model are likely to be misleading. Of these the most important has been stated several times above, but is worth repeating. The balls represent information, not stimulation. The reader must not contemplate the Y tube and decide that two stimuli cannot be dealt with simultaneously. They can if they convey sufficiently little information. Clearly, many reflexes are compatible with one another, and it is likely that simple "voluntary" reactions are equally capable of being carried on simultaneously. It is only with unpracticed reactions involving a choice between several alternatives that we find an interference between two stimuli; but this is a very normal case outside the laboratory. Theories such as that put forward by Welford (38) for the "refractory period" undoubtedly need the qualification that highly probable stimulus sequences may not show these effects (22). But this does not disprove their general value.

A related point is more serious. The length of the immediate memory span is roughly constant whatever the size of the ensemble from which the items are chosen (31): one cannot remember enough binary digits to make the information in immediate memory equal to that stored in memorizing ordinary decimal digits (30). Yet the model makes the limit on memory span dependent on the time taken to pass the items through a limited capacity system, which will in turn depend on the information per item. Perhaps this difficulty may be resolved by suggesting that each possible item represents an extra branch on the stem, and an incoming ball is always recirculated to its appropriate branch. The time taken to withdraw all the balls again would then depend on the time taken by the flap to operate, which is not dependent on the information per item. But this is clearly leading us into complications; for the present we may merely note that the point is an important one about immediate memory but not impossible to handle with a model of this type. A minor caution which should be added is that the model is deterministic while all the experimental results quoted are statistical.

Finally, we have put forward this model as one which may be described by an exactly worded theory which applied also to man. The present paper is directed largely at those who find such a theory unintelligible in its original form, but it should be borne in mind that the theory under test is the abstract one given in the first section. Otherwise the error of identifying the model with the organism may be made, if only to discredit the theory by its obvious absurdity. The formulation given ear-

lier is not complete, but it indicates the way in which the model and the man may be described by the same abstract theory. There should therefore be no excuse for treating the Y tube as anything more than an expository device and a mnemonic for recalling the results of numerous experiments. The writer's freedom from such an error is demonstrated by the fact that he has never built his model in any physical sense.

SUMMARY

A mechanical model is described, to act as an easy introduction to a formal theory of attention and immediate memory in information theory terms. A number of deductions from the theory which agree with experimental results on human beings are given as descriptions of the behavior of the model.

REFERENCES

1. BAHRICK, H. P., NOBLE, M., & FITTS, P. M. Extra task performance as a measure of learning a primary task. *J. exp. Psychol.*, 1954, **48**, 298–302.
2. BERLYNE, D. E. Stimulus intensity and attention in relation to learning theory. *Quart. J. exp. Psychol.*, 1950, **2**, 71–75.
3. BILLS, A. G. Blocking: a new principle in mental fatigue. *Amer. J. Psychol.*, 1931, **43**, 230–245.
4. BROADBENT, D. E. Speaking and listening simultaneously. *J. exp. Psychol.*, 1952, **43**, 267–273.
5. BROADBENT, D. E. Listening to one of two synchronous messages. *J. exp. Psychol.*, 1952, **44**, 51–55.
6. BROADBENT, D. E. Noise, paced performance and vigilance tasks. *Brit. J. Psychol.*, 1953, **44**, 295–303.
7. BROADBENT, D. E. The role of auditory localization in attention and memory span. *J. exp. Psychol.*, 1954, **47**, 191–196.
8. BROADBENT, D. E. Successive responses to simultaneous stimuli. *Quart. J. exp. Psychol.*, 1956, **8**, 145–152.
9. BROADBENT, D. E. Immediate memory and simultaneous stimuli. *Quart. J. exp. Psychol.*, 1957, **9**, 1–11.
10. BROADBENT, D. E. Effects on behaviour from noises of high and low pitch. Brit. Med. Res. Council, *Appl. Psychol. Unit Rep. No. 222*, 1954.
11. BROADBENT, D. E. The concept of capacity and the theory of behaviour. In *Proc. 3rd London Symposium on Information Theory*. London: Butterworths, 1956.
12. BROADBENT, D. E. Listening between and during practised auditory distractions. *Brit. J. Psychol.*, 1956, **47**, 51–60.
13. BROWN, J. The nature of set-to-learn and of intra-material interference in immediate memory. *Quart. J. exp. Psychol.*, 1954, **6**, 141–148.
14. BROWN, J. Immediate memory. Unpublished doctoral thesis, Univer. of Cambridge, 1955.
15. CHERRY, E. C. Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Amer.*, 1953, **25**, 975–979.
16. CONRAD, R. Speed and load stress in a sensori-motor skill. *Brit. J. indust. Med.*, 1951, **8**, 1–7.
17. CONRAD, R. Missed signals in a sensori-motor skill. *J. exp. Psychol.*, 1954, **48**, 1–9.
18. CONRAD, R. Adaptation to time in a sensori-motor skill. *J. exp. Psychol.*, 1955, **49**, 115–121.
19. DEESE, J. Some problems in the theory of vigilance. *Psychol. Rev.*, 1955, **62**, 359–368.
20. DEUTSCH, J. A new type of behaviour theory. *Brit. J. Psychol.*, 1953, **44**, 304–317.
21. DEUTSCH, J. A machine with insight. *Quart. J. exp. Psychol.*, 1954, **6**, 6–11.
22. ELITHORN, A., & LAWRENCE, C. Central inhibition—some refractory observations. *Quart. J. exp. Psychol.*, 1955, **7**, 116–127.
23. GIBSON, J. J., & RAFFEL, G. A technique for investigating retroactive and other inhibitory effects in immediate memory. *J. gen. Psychol.*, 1936, **15**, 107–116.
24. HOVLAND, C. I. Experimental studies in rote learning theory. II. Reminiscence with varying speed of syllable presentation. *J. exp. Psychol.*, 1938, **22**, 338–353.
25. HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp Psychol.*, 1953, **45**, 188–196.
26. KAY, H., & POULTON, E. C. Anticipation in memorizing. *Brit. J. Psychol.*, 1951, **42**, 34–41.

27. MACKWORTH, N. H. Researches on the measurement of human performance. *M.R.C. Spec. Rep. No. 268.* Her Majesty's Stationery Office, 1950.
28. MACKWORTH, J. F., & MACKWORTH, N. H. The overlapping of signals for decisions. *Amer. J. Psychol.*, 1956, **69**, 26–47.
29. MACKAY, D. M. In symposium on cybernetics. *Advanc. Sci.*, 1954, **40**, 402–406.
30. MILLER, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.
31. POLLACK, I. Assimilation of sequentially encoded information. *Amer. J. Psychol.*, 1953, **66**, 421–435.
32. POULTON, E. C. Memorization during recall. *Brit. J. Psychol.*, 1953, **44**, 173–176.
33. POULTON, E. C. Two-channel listening. *J. exp. Psychol.*, 1953, **46**, 91–96.
34. POULTON, E. C. Listening to overlapping calls. *J. exp. Psychol.*, 1956, **52**, 334–339.
35. VINCE, M. A. Rapid response sequences and the psychological refractory period. *Brit. J. Psychol.*, 1949, **40**, 23–40.
36. SPIETH, W., CURTIS, J. F., & WEBSTER, J. C. Responding to one of two simultaneous messages. *J. Acoust. Soc. Amer.*, 1954, **26**, 391–396.
37. WEBSTER, J. C., & THOMPSON, P. O. Some audio considerations in air control towers. *J. Audio. Eng. Soc.*, 1953, **1**, 171–175.
38. WELFORD, A. T. The psychological refractory period and the timing of high-speed performance—a review and a theory. *Brit. J. Psychol.*, 1952, **43**, 2–19.

(Received November 30, 1956)

RIVISTA DI PSICOLOGIA

Publication of the Italian Society of Psychology

Director: PROFESSOR CESARE L. MUSATTI

Co-directors: PROFESSORS ALERTO MARZI, GAETANO KANIZSA, MARIO PONZO

Vol. 50, No. 4: "Jubilee" number in celebration of the fiftieth anniversary of the establishment of the Rivista.

TABLE OF CONTENTS

BARTLETT. **Training Older People: An Experimental Approach to a Social Problem**

BURT. **British Psychology During the Past Fifty Years**

DREVER. **A Note on the Classification of Stimuli**

EYSENCK. **The Questionnaire Measurement of Neurotism and Extraversion**

von FIEANDT. **A Contribution Toward the Problems of Age and Personality**

GEMELLI. **La percezione visiva del movimento**

GOTTSCHALDT. **Handlung und Ausdruck in der Psychologie der Persönlichkeit**

de GROOT. **Über das Denken des Schachspielers**

MEILI. **Geistige Faktoren in den ersten Stadien der Charakterentwicklung**

METZGER. **Ueber Durchsichtigkeits-Erscheinungen**

NUTTIN. **Experimental and Clinical Method in Psychology**

PIAGET. **Centration et décentration perceptives et représentatives**

PIÉRON. **La temperature en Psychophysiologie**

PONZO. **Il fattore della prospettiva in un gruppo di illusioni nelle valutazioni di collettività numeriche**

RAUSCH. **Zur Theorie der sogenannten Vertikaltäuschung**

ROHRACHER. **Zur Theorie der Aufmerksamkeit**

TERSTENJAK. **The Experience of Time and the Question of Subjective Boredom**

WALLON. **Subjectivisme et objectivité en psychologie**

WERNER WAPNER. **Sensory Tonic Field Theory of Perception: Basic Concepts and Experiments**

WELLEK. **Die Entwicklung der Grundannahmen der Psychologie und die Überwindung des Phänomenalismus und Psychologismus**

ZAZZO. **Le bestiaire des enfants**

This "jubilee" number, which will be sent gratis on 1956 subscriptions, will also be sold as a separate issue. Correspondence and subscriptions should be addressed to:

Amministrazione Rivista di Psicologia
Via degli Alfani 37/39
Firenze, Italia

THE BRITISH JOURNAL OF PSYCHOLOGY

Edited by JAMES DREVER

Vol. 48 Part 2 May 1957 20s. net

J. T. BEAUMONT and B. M. FOSS. Individual differences in reacting to delayed auditory feed-back.

H. M. B. HURWITZ and J. CUTTS. Discrimination and operant extinction.

THOMAS MARILL. The psychological refractory phase.

D. G. LEWIS. The normal distribution of intelligence: a critique.

C. B. HINDLEY and FRANK FALKNER. A one-way window.

D. E. BERLYNE. Conflict and choice time.

CYRIL M. FRANKS. Personality factors and the rate of conditioning.

W. T. SINGLETON. An experimental investigation of sewing-machine skill.

J. M. von WRIGHT. A note on the role of 'guidance' in learning.

D. E. BERLYNE. Attention to change, conditioned inhibition (sI_R) and stimulus satiation.

J. R. SMYTHIES. A note on the fallacy of the 'phenomenological fallacy'.

PUBLICATIONS RECENTLY RECEIVED.

OTHER PUBLICATIONS RECEIVED.

The subscription price per volume, payable in advance, is 60s. (or $10) net (post free).

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS

Bentley House, Euston Road, London, N. W. 1

applied psychologists

The Heavy Military Electronic Equipment Department of General Electric has a number of unusual opportunities for psychologists with training and experience in human engineering, industrial psychology and training techniques.

The psychologists will participate in the development and field evaluation of large missile guidance systems with responsibility for:

equipment design
personnel requirements
operations
training

Knowledge of electronic equipment and its operation highly desirable.

A Master's and three years of experience or a Ph.D. essential.

Write in confidence to:

Mr. John S. Brady
Heavy Military Electronic
Equipment Department

GENERAL ELECTRIC

Court Street, Syracuse, N. Y.